On the Pitfalls of Crowdsourcing for Civic Information Management

Konstantinos Pelechrinis, University of Pittsburgh

Abstract
The advancements and proliferation of mobile networking and computing technology has enabled large-scale collaborations between people towards solving specific problems. While crowdsourcing has been utilized to facilitate various research efforts and a fast growing number of businesses utilizes crowdsourcing approaches in their products, it is only recently that the "wisdom of crowds" has seen applications in civic life and urban planning. Whereas the potentials are huge, successful design and deployment is not trivial. The major challenges that local government offices have to face when using crowdsourcing for urban planning operations are: (i) providing incentives for usage from the city-dwellers, while ensuring the quality of information submitted and (ii) providing accessibility to the corresponding platform for the mass of the population. In this article we provide a qualitative discussion on the potential pitfalls of crowdsourcing in managing and exploiting civic information.

Keywords: Urban informatics; Crowdsourcing
Citation: Pelechrinis, K. (2015). On the Pitfalls of Crowdsourcing for Civic Information Management. In *iConference 2015 Proceedings*.
Contact: kpele@pitt.edu

1 Introduction

Currently more than half of the global population resides in cities[1]. Cities have been long considered as the paramount instrument for wealth creation and innovation. However, they have also transformed to the main source of crimes, diseases and pollution. Hence, as massive amounts of people move to urban areas, there is a need for cities to be run more efficiently. Nevertheless, at the same time these massive urban populations make this task much harder and challenging.

Despite the aforementioned conflicting dynamics, many of the operations related to urban planning can be facilitated by appropriate exploitation of the technological advancements in mobile computing and networking as well as from the unprecedented amount of data available from a number of sensors. Ubiquitous mobile computing (e.g., smartphones) has essentially enabled the interconnection of the masses in a previously unseen scale. In turns, this has facilitated the fast propagation of news, trends and practices through the urban space. The potentials to transform the planning and management of megacities towards a bottom-up paradigm are unparalleled and can be realized through *crowdsourcing*.

Crowdsourcing has recently become ubiquitous in the business world, where services utilize the "wisdom of crowds" (e.g., review platforms). Well defined but laborious tasks within research projects (e.g., data labeling) have also found their way to crowdsourcing platforms (e.g., Amazon's Mechanical Turk), while computationally expensive tasks are completed by large crowds who volunteer their machines for the computations[2]. One of the best examples of "urban" crowdsourcing is OpenStreetMaps (`http://www.openstreetmap.org`), where citizens contribute to the creation of open and accessible maps of our cities and beyond. The main idea behind crowdsourcing is the division of a large task, which is practically impossible to be completed by a few people, to a very large number of small components/mini-tasks. Each of these mini-tasks is usually fundamental and does not require specific expertise. As a result, they can be completed with a minimal effort from individuals, and hence, a large group of people can collectively complete the previously impossible task.

As alluded to above, the scale of megacities can cause problems in the completion of urban planning operations. The latter could be a manifestation of either sub-optimal operations caused from the large "solution space" to be explored or of the inability to complete the operations at all. For instance, the spatial expansion of the urban areas in conjunction with their densification and the decline in the local governments' financial resources make it practically impossible for the bureau of building inspection to inspect whether

[1] `http://esa.un.org/unup`
[2] E.g., `http://setiathome.ssl.berkeley.edu`

every small home renovation/construction has acquired the required permits and follows the safety codes. As another example, street clearing under heavy weather conditions (e.g., snowstorms) is crucial to ensure the public safety while driving. Nevertheless, the plow track routes are usually decided a-priori thus, being static and myopic to the current road conditions and unable to adjust dynamically.

Given the common interest of city dwellers in a livable, sustainable and resilient environment, crowdsourcing appears to be a promising approach towards facilitating planning and management operations. In particular, crowdsourcing could enhance information management tasks, such as information acquisition and curation. One important subset of crowdsourcing, of particular importance to citizen science, is *participatory sensing* (Burke et al., 2006). The ubiquitous mobile devices form a network through which individual users can gather analyze and share information using the integrated sensors. However, crowdsourcing techniques is not a panacea! Not everyone is willing to participate, and not all of the city operations can be broken down to simple mini-tasks that can be completed by non-experts. In other words, it is not clear why city dwellers will participate in the process, or how responsibly and unbiased they will act ensuring the quality of the underlying service. Even more importantly, one needs to take into account the digital divide. Does everyone own the required technology (e.g., smartphones etc.) to participate? Will this further increase the divide, empowering even more the "haves" and marginalizing further the "have nots"?

1.1 Current state of civic crowdsourcing

Urban space analysis has gained a lot of attention in the research community recently. In particular, crowdsourcing platforms which aim into quantifying the perception of dwellers for their city have been designed (e.g., (Salesses, Schechtner, & Hildago, 2013; Zambaldi, Pesce, Quercia, & Almeida, 2014; Quercia, Schifanella, & Aiello, 2014; Quercia, O'Hare, & Cramer, 2014)). In such systems (e.g., StreetSeen[3]), usually two random images, typically obtained through Google's Street View, are presented to a user and a perception question is asked (e.g., "Which street looks more safe to walk?"). By analyzing the results researchers can build "perception maps" for the cities but also identify what elements make a street appear safe, beautiful, walkable etc. Cranshaw *et al.* (Cranshaw, Luther, & P.G. Kelley, 2014) further study the best design for a collaborative system that will be able to produce scalable and expressive externalization of such maps. Other research efforts have focused on the design of social systems and applications that will facilitate urban planning or dweller's urban experience. For example, Zambonelli (Zambonelli, 2011) describes how we can crowdsource the maintenance of a city, while Foth *et al.* (Foth, Schroeter, & Anastasiu, 2011) study the potential of crowdsourcing for reporting code violations. Evans-Cowley (Evans-Cowley, 2011) provides a comprehensive overview of a variety of mobile applications that have been built to assist urban planners.

The increased importance of participatory sensing has also led to technical developments on architectures for collecting, transmitting and analyzing the sensed data (e.g., (Koutsopoulos, 2013; Liu, Hui, Branch, Bisdikian, & Yang, 2011; Sherchan et al., 2012; Zaslavsky, Jayaraman, & Krishnaswamy, 2013)). A large number of studies on GIS and urban mapping applications have utilized crowdsourcing (e.g., (Boulos et al., 2011; Kuznetsov, Davis, Cheung, & Paulos, 2011; Reddy et al., 2010)). However, apart from similar research studies a large number of civic crowdsourcing projects have appeared. Especially successful are the "one-time" projects. For example, Map Kibera[4] trained Kibera's residents on how to use GIS mapping technology who later used this technology to create maps for Kibera that did not exist until that moment. For profit applications, such as Foursquare and Yelp, are also civic crowdsourcing platforms that essentially facilitate the urban experience through recommendations and location-based services powered from the crowds/user-base. The whole market of shared economy can also be thought as a form of civic crowdsourcing, since now people can share the infrastructure to support their urban needs with other people potentially leading to more sustainable cities (e.g., parking sharing - Park Circa, taxi sharing service - Lyft, Uber - etc.).

Despite the recent surge in the field, there are critical pitfalls possible. Therefore, our objective in this article is to enlist these possible shortcomings, but also to showcase the potentials of crowdsourcing and participatory sensing in civic operations. If we are to adapt this design for facilitating the operations of the 21st century city, we need to make sure that this will improve the quality of urban life in a holistic and not superficial way. While solving the digital divide problem is highly complex and we are not even qualified to provide a solution, we will make suggestions on how to avoid some of the possible pitfalls focusing first on

[3] http://streetseen.osu.edu
[4] http://mapkibera.org

specific examples on city operations and later generalizing them. This work can be viewed as a preliminary study required for understanding the power but most importantly the potential pitfalls of crowdsourcing in civic systems, which will in turn ensure their effective design.

Roadmap: The rest of the paper is organized as follows. Section 2 gives an overview of crucial obstacles that can be faced by a civic crowdsourcing system. Sections 3 and 4 elaborate on ways that crowdsourcing can be exploited by cities through two detailed examples, while we provide our concluding remarks in Section 5. We also briefly touch upon issues outside of the scope of our article but important in the general area of civic crowdsourcing (e.g., privacy).

2 What Can Go Wrong with Civic Crowdsourcing?

The homophilous patterns of social connections predict that the majority of our close social circle will exhibit similar ideas, behavior and mindset with us on a variety of issues (McPherson, Smith-Lovin, & Cook, 2001). This phenomenon can many times create the illusion to people that everyone else - outside this tight social circle - thinks and acts in the same way with us. This illusion poses a significant risk on the success of every crowdsourced civic operation and is being realized through two main instantiations:

- **Incentive/Will illusion** - if I am willing to participate "for the good of my city", everyone else in the city is also willing to.
- **Technology access illusion** - if me and my social circle has access to the required technology this must be true for everyone else too.

However, the truth is very different. Not everyone has the incentive to participate. This is not because they do not want to help their city, but because things we take for granted are not such for others. Imagine a dweller who has to work multiple jobs to provide the absolutely necessary for his family. The latter is his/her constant everyday goal and no one should expect from him/her to participate in any volunteering urban crowdsourcing and sensing effort. Similar arguments support the technology access illusion. Not everyone has a smartphone, or even a cell phone. Not everyone is connected to the Internet and the Web. The latter is especially important not only for civic crowdsourcing but also for the delivery of the next generation urban services. Services delivered through cutting edge technology platforms will exclude people that have no access to them, with the latter being most probably the ones that need them the most.

While the above problems might not be important for the design of a commercial application/service based on the crowds - after all a business is only interested to a specific part of the population - they are essential in relation to urban services. The latter affect the urban population as a whole and in general, ignoring the aforementioned risks can lead to system architectures that are suboptimal and not inclusive. In the following two sections we will present some guidelines on how we can improve the chances of a successful crowdsourced civic application. In each section, we will describe a specific example in details and distill the key elements that can serve as general guidelines.

3 Creating Incentives

In order for city planners to be able to develop and integrate urban intelligence with the city operations it is crucial for them to collect the necessary data that will inform their actions. Crowdsourcing can definitely help by requiring every participant to provide a small piece of this information (e.g., public transportation usage, places visited during the course of the day etc.). However, as alluded to above there are no clear incentives for people to participate in such information collection and sharing process. Even if they choose to do so, the information provided can be partial, conflicting and in general of low quality and untrustworthy. After all city-dwellers, especially in large cities, are kept busy enough by the fast pace of urban life to think about participating in efforts from which they have no direct, tangible, benefit.

One solution to this participation incentive problem is to *hide* the crowdsourcing mini-task behind a front-end service. Simply put, the local government can provide a convenience - necessity (preferably) or otherwise - service to its dwellers, which will require to indirectly and accurately share the complete

information of interest. In the rest of this section, we will further explain the context of "indirect crowdsourcing" by using as an example the scenario of a mobile parking payments application.

Mobile Parking Payments for Parking Data Collection

A study of major downtown areas has shown that, on average, 30% of congestion traffic is due to drivers cruising for a free parking spot (Shoup, 2006).Understanding the parking behavior and demand of city dwellers is thus crucial for better parking management. This in turn can help reduce congestion, gas emissions and even improve the customer flow to local businesses. As another example, knowledge about patterns of street parking can lead to educated decisions related with the creation of new parking infrastructure to developing neighborhoods. Furthermore, dynamic pricing schemes based on historical patterns can be developed, which can achieve the optimal utilization of the public parking infrastructure (i.e., operating at a utilization just below capacity) and also potentially increase the revenues of cash-strapped local governments. In turns these additional financial resources can fund crucial projects for urban development.

However, a strong requirement for all these is access to detailed information about the parking behavior of the dwellers. In particular, urban planners will be interested in parking data, that is, information about the number of cars parked on the streets at every point of time. Nevertheless, such data cannot be easily collected, especially when the traditional coin park meters are used! An optimistic city planner could envision building a crowdsourcing application through which dwellers would inform the local authorities when, where and for how long they parked. Alternatively a participatory sensing application could potentially automatically detect through spatial trajectory mining and mode of travel detection, when the dweller parked without requiring direct input. However, as aforementioned this would have been a rather *naive* expectation, since people are too busy to spend the extra time to (directly or indirectly) complete this task even when they have the best intentions to do it[5]. Even if some of them did it, the data would be extremely sparse and for all practical purposes not useful. A potential solution to this problem though is to have citizens share this information with the city as part of a necessary action that **has to** be completed while utilizing the public parking service. In this scenario this action is paying for the parking spot.

In particular, by providing a system that instead of explicitly asking for the parking event information (i.e., the *naive* crowdsourcing application) it rather facilitates it, urban planners will be able to dramatically increase the participation rate and avoid the incentive illusion altogether. In this specific scenario, a payment system with logging capabilities would achieve the goal. There are many options for that and one of them is a mobile payment system. However, we need to be careful because such a solution while avoiding the incentive illusion it does not do so for the technology access one (e.g., not all people have smartphones that will support the paying application or a cell phone where you could potentially use a toll-free text for the payment or even a credit card that is required for this kind of payments). An approach that has been put in practice in the City of Pittsburgh during the last 2 years is parking payment kiosks (see Figure 1). These are placed at every block where paid public parking is enforced and dwellers can pay either with credit card or coins. Every transactions is logged and information is sent back to the parking authority in the tuple format: `<location, date, time, pay_type, pay_time_interval, parking_rate>`, which is exactly the type of information needed for studying the parking behavior of the city-dwellers. Note here also that a (positive) side effect of parking enforcement is the fact that the provided information is very accurate as well[6].

Figure 1: A Pittsburgh parking kiosk by Cale America Inc.

Suggested principle

The key element of the above approach, which overcomes the participation incentive in crowdsourced systems, is the **vital nature of the crowdsourcing mini-task**. The applicability of the above service-based crowdsourcing paradigm is not limited to the parking scenario described. For example, one could envision a similar architecture for public transportation fare payments. This would essentially allow users to pay through smart-devices equipped with logging capabilities (e.g., smart cards provided by the corresponding

[5]Even an automated participatory sensing solution will require from dwellers to spend time download and setup the application.

[6]While information about dwellers that park illegally will not be present, depending on the type of data analysis at hand, we can account for such incidents by utilizing parking citation datasets.

port authority etc.). Embarkment would be possible after payment (e.g., the payment activates the door) and disembarkation would require another "scanning" of the equipment[7]. This provides convenience to the passengers (e.g., no need for exact fare changes) but it also enables transportation authorities to log fine grained statistics for the public transport usage. Until now such estimates have been based on statistical models that make use of coarse-grain information (e.g., (Pulugurtha & Agurla, 2012)).

4 Avoiding the Digital Divide

One of the hardest pitfalls to overcome when designing a civic crowdsourced system is that of digital divide. The question is how to ensure that any application designed will not exclude part of the population that does not have access to the required technology. As it is evident, this is an extremely complex problem to solve, and while a solution cannot be offered within the limits of an article, we will provide directions and suggestions that could lead to solutions that avoid this pitfall.

When one thinks of a crowdsourcing solution he/she usually envisions the participation of a large amount of people/users. However, there are tasks where such a participation is not necessary even though it has the potential to slightly increase the quality of the solution offered. Such tasks typically require geographic coverage rather than population coverage. One such example is air quality monitoring. In this scenario, the responsible authorities do not attempt to cover each and every person in the city measuring the quality of the air they breathe, but rather they distribute sensors across the city ensuring that they cover the maximum possible area. Putting the above principle in the context of civic crowdsourcing, we will examine a toy example of a road condition reporting that can find important applications related to public safety.

Road Condition Reporting

Keeping the city streets clean and safe when there is severe weather (e.g., snowstorms) is of utmost importance for any local government. Typically plow tracks follow pre-defined snow emergency routes, with the later being static. Despite the rationale behind the choice of these routes (e.g., prioritizing primary over secondary roads), many times this approach can be inefficient and/or not very informative to the residents. However, there is currently no way for the responsible authorities to have access to dynamic information of road conditions. One possibility is to allow the residents themselves to report on the conditions of the streets. This would create the ability to generate a map layer (e.g., similar to the one depicted in Figure 2) where this information is presented to city-dwellers so as to know which streets are safe to drive. The city could also utilize this information to further adjust the street cleaning schedule of the plow tracks (e.g., primary roads that are already in good condition might not need to be treated with salt again and the resources can be allocated to secondary roads). Key to such an application is the fact that in order for this service to be offered through crowdsourcing not every single resident needs to report. Only geographic coverage of the city is required and hence, even "one report per street" might be enough. Of course, the more people that make the effort to act as urban sensors, the more accurate and timely the information will be. However, it is not a strict requirement and therefore, crowdsourcing can *bypass* the digital divide problem in similar situations.

Suggested Principle

The overarching principle in the above discussion is the requirement for **geographic coverage**. When geographic coverage is required, crowdsourcing *may* bypass the digital divide problem and provide a viable solution. Emphasis should be given though on the fact that this is not always a sufficient condition for a successful service. When geographic coverage is required, we also need to make sure that the technology gap of the city-dwellers is not mapped on a geographic segregation as well. In this case, we can end up with areas being over and under represented and this can lead to different areas enjoying different quality of service. Therefore, while a service that requires geographic coverage is an ideal candidate for a crowdsourced solution, attention still needs to be taken in order for all the residents to be treated in an equal manner (e.g., data analytics methods that account for similar biases).

[7]Note here that, such systems currently exist in various cities, mainly for their subway systems.

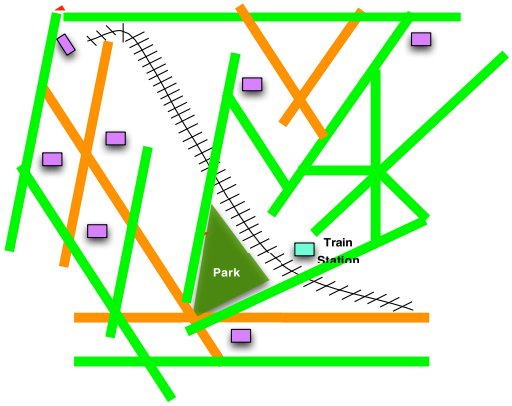

Figure 2: An illustrative example of the output of a road condition reporting application where street segments are color-coded based on their driving risk. Discrete levels of safety are used in this example (red: dangerous, orange: medium safe, green: safe).

5 Discussion and Conclusions

In this article we have discussed the potential pitfalls of civic crowdsourcing. While the possible benefits are well understood and documented, little attention has been given to what can go wrong when cities start basing their decisions on the wisdom of the crowds. We specifically focus on two critical problems that need to be considered, namely, the incentive problem and the digital divide. We also provide some general principles that could bypass - not solve - these problems.

However, we are still far from being close to developing civic applications that offer city services equally to every citizen. Even when dwellers have enough incentives and access to the technology to participate in urban sensing, privacy concerns will come up. Even in the case of indirect crowdsourcing (Section 3) the way data are used from authorities need to ensure privacy and made known to the public. Equally as important, efficient ways to integrate the information collected with the city services is required. How will the information about road conditions will be integrated in the city's workflow? Reporting is only the beginning and the city is responsible for dispatching the units and finally delivering the service, which requires careful integration with the operational workflow. This article is not implying that civic crowdsourcing is condemned. On the contrary, we believe that the potentials are unlimited. However, a stepping stone towards fully realizing them is understanding the latent dangers and difficulties. We firmly believe that with the points raised in this article we will stimulate further research and discussion on the topic, which we also seek to undertake as part of our future research agenda.

References

Boulos, M., Resch, B., Crowley, D., Breslin, J., Sohn, G., Burtner, R., ... Chuang, K.-Y. (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, ogc standards and application examples. *International Journal of Health Geographics*, *10*(1). Retrieved from http://dx.doi.org/10.1186/1476-072X-10-67 doi: 10.1186/1476-072X-10-67

Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., & Srivastava, M. B. (2006). Participatory sensing. In *International workshop on world-sensor-web.*

Cranshaw, J., Luther, K., & P.G. Kelley, N. S. (2014). Curated city: Capturing individual city guides through social curation. In *Acm chi.*

Evans-Cowley, J. (2011). There's an app for that: Mobile applications for urban planning. In *Social science research network.*

Foth, M., Schroeter, R., & Anastasiu, I. (2011). Fixing the city one photo at a time: mobile logging of maintenance requests. In *Acm ozchi.*

Koutsopoulos, I. (2013, April). Optimal incentive-driven design of participatory sensing systems. In *Infocom, 2013 proceedings ieee* (p. 1402-1410).

Kuznetsov, S., Davis, G., Cheung, J., & Paulos, E. (2011). Ceci n'est pas une pipe bombe: Authoring urban landscapes with air quality sensors. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 2375–2384). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1978942.1979290 doi: 10.1145/1978942.1979290

Liu, C., Hui, P., Branch, J., Bisdikian, C., & Yang, B. (2011, June). Efficient network management for context-aware participatory sensing. In *Sensor, mesh and ad hoc communications and networks (secon), 2011 8th annual ieee communications society conference on* (p. 116-124).

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. In *Annual review of sociology, 27:415-44.*

Pulugurtha, S., & Agurla, M. (2012). Assessment of models to estimate bus-stop level transit ridership using spatial modeling methods. *Journal of Public Transportation*, *15*(1).

Quercia, D., O'Hare, N., & Cramer, H. (2014). Aesthetic capital: What makes london look beautiful, quiet and happy? In *Aaai cscw.*

Quercia, D., Schifanella, R., & Aiello, L. (2014). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Acm hypertext.*

Reddy, S., Shilton, K., Denisov, G., Cenizal, C., Estrin, D., & Srivastava, M. (2010). Biketastic: Sensing and mapping for better biking. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1817–1820). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1753326.1753598 doi: 10.1145/1753326.1753598

Salesses, P., Schechtner, K., & Hildago, C. (2013). The collaborative image of the city: Mapping the inequality of urban perception. In *Plos one 8(7): e68400, doi:10.1371/journal.pone.0068400.*

Sherchan, W., Jayaraman, P. P., Krishnaswamy, S., Zaslavsky, A. B., Loke, S. W., & Sinha, A. (2012). Using on-the-move mining for mobile crowdsensing. In *Mdm'12* (p. 115-124).

Shoup, D. (2006). Cruising for parking. *Transport Policy*, *13*(6), 479–486.

Zambaldi, V., Pesce, J., Quercia, D., & Almeida, V. (2014). Lightweight contextual ranking of city pictures: Urban sociology to the rescue. In *Aaai icwsm.*

Zambonelli, F. (2011). Pervasive urban crowdsourcing: Visions and challenges. In *Ieee percom workshops.*

Zaslavsky, A. B., Jayaraman, P. P., & Krishnaswamy, S. (2013). Sharelikescrowd: Mobile analytics for participatory sensing and crowd-sourcing applications. In *Icde workshops.* IEEE Computer Society.