NIKHIL RAIZADA  
218541102

Submitted To: Dr Niroshinie Fernando (Supervisor)

Team 2

Assignment 4 - Minor Thesis   
SIT792-Minor Thesis

OPPURTUNITIES AND CHALLENGES IN CROWD ENABLED COMPUTING

Contents

Abstract………………………………………………….….…….….3

Introduction………………………………...……………………...…3

Background and Motivation………………………………….……...5

Key Related Research Review …………….…………………...…...8

Research Methodology.…………….……………………. .…..…....15

Problem Analysis……………………………………………………15

Result………………………………………………………………...17

Conclusion………….……………….………………..……………..23

References………………………………………………….…….….25

Abstract

This thesis outlines different projects on crowd-enabled computing focusing on mobile phone and the applications which are being used for crowdsourcing to help a city to overcome the real-life problems and modify it into a smart city. Crowd computing with mobile sensing have a great potential and can produce huge amount of processing power and data if implemented efficiently and adopted by masses. This study involves solving problems which today’s urban environment suffers like traffic congestion and accidents, crime, pollution and many other issues. We have analyzed many research papers on crowd enabled systems and application which aims to solve the similar problems that we are addressing in this paper. Crowd sourced computing can provide solutions to many problems which cannot be solved by conventional computing as this distributed system approaches the issues by contributing from individuals. Groups and communities.

Introduction

Crowdsourcing is the model of involving group of people for a common goal. Main motive behind crowdsourcing is innovation, efficiency or problem solving. It is mostly done by social media or web applications [25]. It can be helpful for any organization as it provides solution for the problems which we face in our day to day life. Since, the usage of smart phones have become handy. It has played an important role in making crowdsourcing successful. The emergence and evolution of smartphones played a vital role in increasing the communication efficiency by overcoming the barrier of transferring information over long distance in almost real time manner. With the introduction of smartphone in the early 1990's, there is a paradigm shift on how the people communicate and transfer personal data, over last couple of decades. The cumulative median of **percentage of population using the smartphones of different countries increases to 54% in 2015 from 45% in 2013** [1]. Smartphones Network is one of the most prominent worldwide technological architecture comprises of millions of nodes where a smartphone depicts a single node and each smartphone is connected to every other smartphone directly or indirectly. 1700 millions mobile phones are projected to be sold in year 2020 with around 120-184 gigabytes of data transferred from each device [2]. If data from a single smart phone is analyzed, we can find out behavioral and communication patterns, likes and dislikes and many other physical aspects like location and movement of the user.

Source: Statista

This distributed rich network of smart portable devices in which million bytes of data is transferring each second can be used as a task engine or a media to perform operations which are very difficult or impossible for a conventional single point system. **Crowd Sourcing** [3] when combined with smart phone distributed network can be used where computing power or data from multiple systems or specifically mobile phones is aggregated to achieve a single goal. **This method of distributed smart phone of individual users cooperating simultaneously with one another to produce an output is termed as Crowd Computing** [4]. Although a lot of work has been done in development of this field and a lot of applications are powered by Crowd Computing, we are still far from exploiting its full potential. One of the best and popular example of crowd computing especially where use of smart phones played a huge role is **Open Street Maps** [5], which is an open licensed digital world map built by collecting the geospatial data shared by people and aggregated to the cloud and data is available free of cost Unlike other mapping services like Google Maps.

Aggregator

Computing power/data from crowd

Data preprocessing, analysis and information extraction

P

Query

Data representation portal

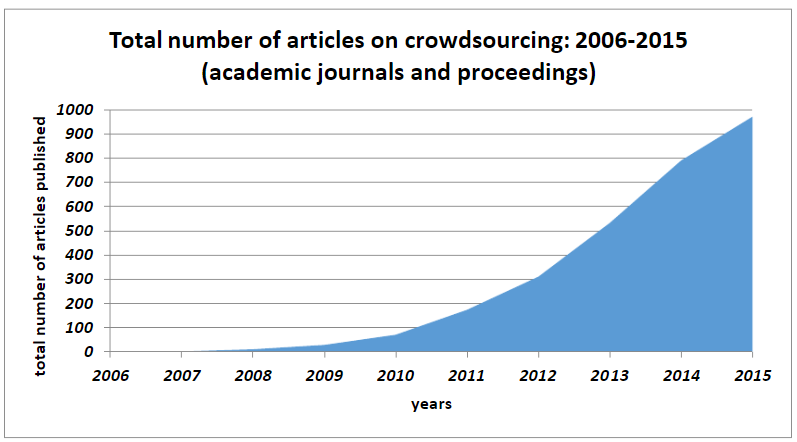
Result

Processed Data Storage

Basic Schematic representation of crowd-enabled computing

Background and Motivation

Crowdsourcing, that is trending all over the world today, is not a term that originated overnight. It is related long back to 1714[26] when British government, seeking innovations, announced £20,000 for the people to come up the best idea, that must be valid and dependable to measure longitudes at sea and eventually it was won by John Harrison. In June 2006, John Howe coined the term Crowdsourcing in his article “Rise of Crowdsourcing.”[26].There are many more examples to relate Crowdsourcing to past. Although it has been around for a long time, but now Crowdsourcing is getting the world by storm and finally is getting along. Since then, Crowdsourcing has helped in a lot of ways, from greatest inventions to biggest brands present today. Since the development of internet 2.0, many organizations has been setup that provides a dedicated platform that help businesses to find solutions to their problems and give them the best advice. Social media platforms are also among the best platform to look when gathering ideas for new products and services. In the below image look at the increase in number of articles published on Crowdsourcing over the decade.[30]



Our study outlines the history of some studies and cases underwent in the past helping in environment issues. To our knowledge till date, Crowdsourcing has entered in various areas of utilization for example content writing, business growth, public contest, real-time alerts, Wikipedia etc. Crowdsourcing is so attractive because without the help of people the project can burnt a hole in the pocket, as renting workspaces and building infrastructure to accommodate people to make a database for a dictionary is less effective than Crowdsourcing the database. The more the number of people of a project the more the database accrued at minimal costs. Crowdsourcing also manages to merge the work environment into friendly conditions. It also gives the advantage to businesses to put their requirements on crowdsourcing sites to get that done at an affordable price and at greater speed in real-time.

There are many examples currently in the world that are using crowdsourcing to lead the world with new innovations and cutting-edge ideas just from its fans and customers.[27]

* DeWalt: DeWalt’s community insights provides a way for the company’s customers and well-wishers to contribute to the product development cycle and resulted in a popular innovation of cordless hammer drills. They replace the expensive and time-consuming methods of traditional customer research with real-time access to more than 10,000 active consumers. They test, launches and then demonstrates the value of premium products. Members of the community said,” they feel appreciated and are comfortable playing an active role in the design and product development process but this could only happen because of the engagement of DeWalt with the active community on a regular basis. This has saved them over $1 million in study costs in 2016 and around $6 million since establishing the DeWalt insight panel.
* Unilever: Launched in 2010, its open innovation platform asks out to the public for solutions for its technical problems on things like smart packaging and cooling systems. Now, the company receives thousands of responses each year. They also make partnerships with established businesses, individual inventors and anyone else with a practical innovation than can help them meet their challenges.[28]

* DHL: It’s innovation centers provides a way to the customers to contribute their ideas for improved systems and technology, which includes drone delivery technology and remote sensing and navigation control tools.
* Airbeam: It is a mobile device produced by HabitatMap that can be worn or installed indoors or outdoors to keep a tab on pollution levels in our surroundings around the clock. It measures a particular matter by drawing the concentration and size of particles using a ray of light. This device communicated measurements to the Android devices every second via Bluetooth or over the internet. Its data is then organised and crowdsourced on their website, which allow them to know the areas that are highest and lowest polluted and share it with the community. This is how crowdsourcing is helpful in environmental monitoring.
* SciStarter: This is an online community that is dedicated in improving the citizen science experiment for project managers and students. Over hundred thousand projects and events are searchable by location, topics and age levels. After joining the organisation members can track their contributions and provide valuable feedback.

These are among the many other organizations that values crowdsourcing.

**Motivation**

There are few good examples of Crowdsourcing which gratified in various ways and it is helping us to make our world a better place with time every day. These initiatives were done either in a city or an area where the situation lies which also explains that crowdsourcing can be useful in finding ways and answers for every situation. Though there were many limitations with every study and not every research was explored up to the optimal level, but with this start of knowledge we can exploit this domain more over the time with more resources and larger community.

* Waze platform [32]-[34]: This is a map-based platform where people get to connect with each other in real-time scenario. This study was conducted in Stockholm, Sweden.Waze help them to monitor road traffic incidents at real time through a map-based application as well as a website. This mobile friendly platform works smoothly to alert the other commuters on the road as well the authorities to keep a check on the incidents happening around the city in real-time. Basic principal of its working is crowdsourcing, where people voluntarily has joined this initiative to give real time alerts from different places situated in different parts of the city and update the situation through the Mobile friendly application. The main disadvantage that they had found was the authenticity, latency and various other factors that have been discussed later in this project. But except some of these factors it has been really a very good initiative of crowdsourcing in environmental monitoring and has various benefits later described.
* Contest based water quality monitoring scheme [29]: This paper was about the contamination of public water with impurities and the most stressed impurity was Lead. There were many previous exercises done to check the impurity level but it was never done with tap waters at private properties. So, as user level sampling was not done in a proper way. So, over the time, research body comes with an idea of crowdsourcing and engaging public in environmental monitoring to address the drinking water crisis. So, a contest-based water quality monitoring scheme was launched for the consumer to check water quality of the tap water at their property. Through this crowdsourcing scheme they collected tap water and conducted analysis of it to better understand the situation. This scheme was carried at New Hampshire in 2018-19 which had a population of around 30,797 people and 12,953 households at that time. The scheme was cash incentive of worth $200 for the one who collected the most samples from different locations to boost the level of participants.800 packets were distributed to people to bring them with water but with an 18% return rate only 142 packets came and were tested for lead concentrations over a period of 26 days. These samples were taken from various taps around the city and over 68% had lead concentration below 1mg/L and 3% had severe lead concentration and some had extreme level of concentration that were taken from wells. So, this study also tells us that lead exposure in private systems remain a large data gap which makes a barrier for protection of public health.

Key Related Research Review

By using the basic principles and architecture of crowd mobile sensing to solve the problems of urban city environment like

* + - Traffic congestion and road accidents [6],[7]
    - Air pollution [8]
    - Monitoring and tracing criminal activities [9]
    - Disaster alerts like earthquake [10]
    - Hotel, restaurant or parking reservation [11]
    - City information guide for new citizens or passersby,

we can transform a standard city to a smart one [12]-[14], where the problems faced by citizens will be solved by citizens collectively with the help of crowd computing.

The transport medium inside a large city plays a significant part in efficient working of a city. A city with well-planned road and traffic system will definitely have more fluid inter-city movement than a city with poor road management. Traffic congestion is termed as a condition in which drivers drive at a comparatively slower speeds than the permissible limit due to various constraints like extensive users on road, bottlenecking/funneling, inappropriate lane change and mishaps [15]. **Traffic congestion can have effect on economy, increase stress for drivers, road-rage and increase environmental pollution** [15]. Cities can adopt intelligent measures to counter the traffic congestion and reduce its ill effect on its citizens and a lot of work has been done in this regard like a **Social Network of vehicles** powered by IOV ( Internet of Vehicles )[6] and crowd sensing which can help with smooth traffic movement and predict traffic and congestion elevations in smart cities. Other similar example is a mobile app based on IOT and crowd computing for the Aarhus city of Denmark which can be used for traffic monitoring [16].

Congestion and road accidents can also reduce by deploying fleet of smart autonomous vehicles and one such crowd sourced project catered by renowned programmer George Hotz [17] is **Comma AI’s Open Pilot** [18] **which convert a non-autonomous vehicle into an autonomous one with the help of mobile phone combined with car sensors and their state of art software built on driving data provided by people** and other technologies such as computer vision, ML and Deep learning. All of the above discussed applications utilizes technologies such as crowd computing, mobile sensing and web of things trying to solve well known problem of traffic congestion and helping to reconstruct a city to smart city.

Elevated pollution levels are also a concern for large cities with huge population, so monitoring and lowering of pollution is also on high priority for a smart city for the well-being of its citizens. Some crowd computing and mobile sensing intelligent solutions can help the city council like a platform called **AirSense**[8] which with the help of a smart phone paired with AQMD ( Air quality monitoring device - built on Arduino pro mini) collect data about air pollution and store it to cloud for analysis and representation purpose. **NoiseTube**[19] is project just like AirSense monitors and analyze the pollution level for noise instead of air, although it doesn’t require any external module and collects information from mobile microphone sensor of the person. It is also crowd sourced and shows the area where is escalated noise level by pairing the noise data with geo coordinate information from people’s smart phone GPS module.

Many other crowd computing research aimed at developing smart city are: -

1. **Transafe** - Transafe is a proposed crowd sourced platform which collects data from crowd, mobile sensor and input from law enforcement services regarding suspicious and criminal activities, aggregates it and after analyzing represent the output to crowd and government officials [9].
2. **ParkSense** - It is a smart automobile parking management system which gives almost real time information about parking availability in nearby parking spot by getting geo coordinate data from its users and also getting data from wifi beacons put on the parking spaces. The system takes account of all the in-out traffic from a parking space by monitoring wifi signal strength of individual smart phone and analyze if a parking space is empty or filled [20].
3. **BikeNet** - It is mobile sensing approach where multiple sensors include smart phone are put on cyclist’s bicycle to collect information about the riding experience and give valuable information such as performance, environmental data such as pollution level and terrain semantics and other cyclist information [21].

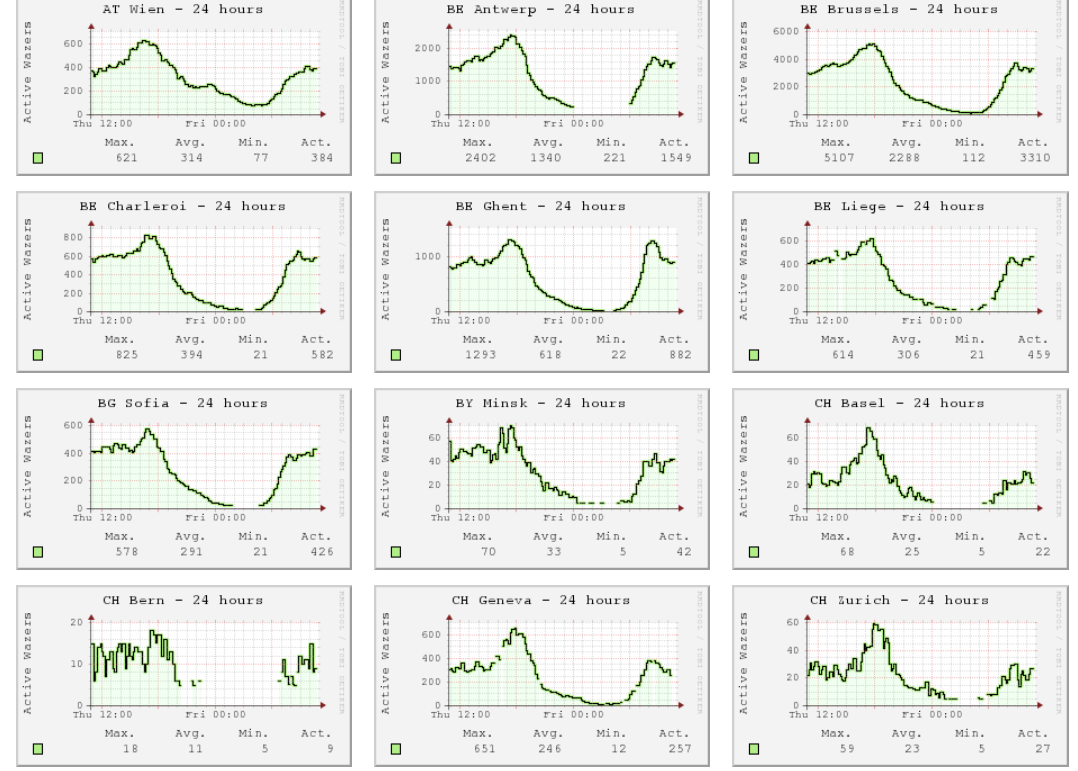
Though the crowd enabled projects and systems discussed above have intelligent concept and works on the principles of cloud computing, some of them may fail to implement in real life scenarios. Some of the general and common weakness or hurdles in project discussed are:

* + **Lack of scalability factor and robustness** - Many of the systems discussed above have high probability of failure when scaled to large masses or have need to redesign the architecture to scale properly. For e.g. To implement and execute the idea of social network of vehicles, a huge amount of investment for both capital and time is needed to create hardware and software for communication between vehicles. The idea of ParkSense depend highly on wifi beacons and fitting every parking spot with this technology is a tedious task.
  + **Unwillingness of crowd to participate** - The crowd plays the most import role in crowd computing. Without active participation of people, one cannot build a crowd enabled system.

**Privacy of user is compromised** - As the crowd computing systems are built upon data and information provided by smartphone user, if strict measures not taken, it can highly compromise the privacy of user especially in systems like OpenPilot, Transafe and NoiseTube.

Following is the case study which was a better example of crowd-sourcing happening around the world and mostly in Europe.

Unforeseeable events, such as accidents and vehicle breakdowns, can lead to traffic congestions which may cause more disasters. So, the early observation and elimination of any such incidents is very important. Most countries in the world faces the challenge of getting an accurate picture of traffic conditions in real time. Mostly conventional methods are used- like traffic cameras or phone calls received at the time of incident. However, these methods are not optimal. Whereas, in 21st century, there are various automated traffic monitoring devices and techniques, but there are still challenges in obtaining the real situation of the traffic conditions. So, the alternative way of traffic monitoring except form machine learning is, human based learning or called as crowdsourcing. This will not only help to save us time but sources too. These types of solutions are best in the developing countries, where the installation of high-end technology is not appropriate. So to make all this happen, there is a navigation application, which works on the crowd-sourcing principle, called WAZE [31]. In 2014, Waze started a collaboration, called the Connected Citizens Program (CCP), with traffic agencies and road operators to exchange real-time traffic information. It is an application that connects the drivers in a community, who share real-time information about the road traffic conditions and give live updates about any disasters that they see while driving ahead. This helps fellow drivers to save themselves from upcoming hazards and advise them various other routes that they can follow to avoid the upcoming situation. It also learns from user liking and disliking, route one prefers the most etc. The study area was limited to Sweden before, but now it has been expanding all over the globe [32]. The major work of this technology is to collect sufficient information and integrate it with other available sources to form an aggregated, easier to comprehend data. This reduces the cost and captures details from the angle, which is unable to get form sensors or camera, which is also a great advantage of Crowdsourcing in data collection. The value and efficiency of this application depends upon the Number of users who provides legitimate traffic information which leads to more data and gives a boost to crowdsourcing. Apart from that, very important members of the Waze community are the map editors who keep the map up to date by editing road segments and marking road closures voluntarily. They collect the information about road closures from the websites and services of road operators and edit the map manually.

****

**Figure 1. No of active crowd providing critical information in various capitals across Europe [38].**

**Figure 2. The alert in the Traffic View Tool[31]**



Weekdays

Weekends

50

45

40

35

30

25

20

15

10

5

0

0

2

4

6

8 10 12 14 16 18 20 22

TIME (HOURS)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

AVERAGE NUMBER OF REPROTS

**Figure 3. Average number number of alerts throughout a day in Stockholm. Weekdays and Weekends are separated.[31]**



**Figure 4. Traffic flow along custom-specified routes and the map in editing practice mode [editing tool for crowd][34]**

**Challenges in collecting traffic information.**

Major challenges faced in thid study is to get a relaible real-time image about the traffic situation were there is a delay in reproting the incident which leads to fatal consequences due to lack of instant cure at the sight. The problem is thoroughly presented in (Rudmark & Arnestrand, 2018). However, there are various other problems that people get that has to be considered[31][33].

• A map error somewhere along the preferred route.

• A lack of correct traffic data for that route.

• A lack of correct traffic data on the route it is trying to send to.

• A data network connection would be present all, or at least most of the times, Without an internet connection, we won't be able to locate or navigate a route.

• Delay in traffic incident reporting where automated incident detection is not installed.

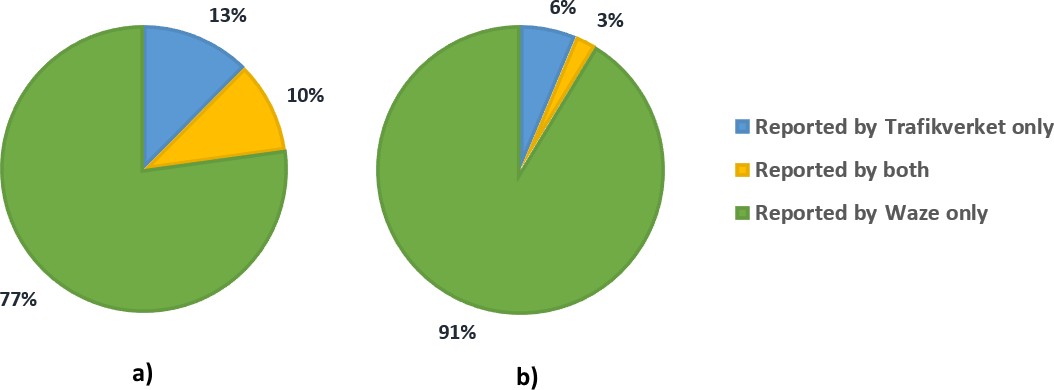
• Inefficient ways of informing the travelers about traffic conditions.

**Result and Analysis**

After the unwanted information was weed out, the descriptive analysis of Waze data was conducted. It was focused on the Waze alerts and only specific alert types. Firstly, it was examined the spread of the alerts on the grass root level to identify regions where the use of Waze is prominent. From the number of inputs from users of an area, we could infer to the extent of the usage of Waze in that area. After this step, the study area was limited to the region where the popularity of Waze is approximately uniform and local deviation would not influence the result of the analysis. The latter, finer analysis, looked at the number of the alerts per road type to get an understanding of the distribution of the alerts in a more detailed way[35]. The severity of traffic incidents have different perspective [36]. This study focused on the impact of the incidents made on the roads for commutors. This study does not provide any information about the severity of the alerts in the data feed. Unfortunately, there was no severity information about alerts, which did not allow the construction of a model for severity prediction which is still a big flaw about crowd-sourcing.

**Conclusion**

This study explored the possibilities and challenges of utilizing crowdsourcing technologies in traffic incidents . In form of a case study, crowdsourceded traffic information obtained from Waze was analyzed and compared with the existing traffic information (trafikveret)in Stockholm. In this topic the insights gained during the study are summarized with the aim to present the potential benefits of **Crowd-sourcing**.



**Incidents detected by Crowdsourcing through waze and Trafikverket separately for big (a) and small (b) roads.**

Hence, by crowd-sourcing the information to Waze, the emergency department in various countries can take optimal steps and make their resources fully for the people.

The expected benefits that we can get are-

* To get traffic information from the areas that are not covered with advance automated detection systems.
* To get additional information about the incidents reported through other means.
* To get the information at the earliest, so that requried steps can be taken to solve the issue.
* To get information for more incidents, so that departments can improve their efficiency in work.
* To get additional information (e.g. potholes, traffic light faults etc.) from the users.
* Moreover, foreign travelers using this can get a more comprehensible prespective of the situations , especially in case of road closures.

In this case study still this domain of technology is not being exploited yet, like various others, by biggest companies in the world. So, in many more years to come we can see a huge leap of development in this domain

Research methodology

Firstly, we started exploring the concepts of crowdsourcing in general where we went through different domains (Environment monitoring, Transportation and traffic planning, Mobile social recommendation, Healthcare, Public safety) in crowdsourcing. Then according to our research findings, we narrowed down our research specifically on environment monitoring and traffic problems in the city. Since, our idea was to take the data from user’s mobile phone in crowdsourcing. Therefore, we found the research papers where crowdsourcing is done by user’s mobile phone and the applications which are being used for crowdsourcing. Then, we identified the research questions in our area of research. Further research was focused on those research questions where we gathered the data related to our research question. After that we analyzed the research papers and organized our knowledge which we gained from the papers. We searched about the methodologies that are used in the different research papers to overcome the similar problems that we are addressing in our thesis. By comparing the different methods and techniques from our study we have tried to give the optimal solution of our research problems.

Problem Analysis:

In our Study, our focus revolves around the challenges that are faced in crowdsourcing in various projects. We came across so many wonderful projects and studies that focused on helping mankind with resources and time. But, in this digital world where technology helps us to connect and transfer information that is helpful to grow the deed, we came across so many challenges, these studies faced. We worked on various studies to find all the common challenges they faced and, tried to get some results that best suited for their challenges. So in this part we are going to talk about the key challenges and what questions we came across while studying them.

**Research Problems:**

The key challenges that we came across are:

* **Availability of resources [42]**: One of the basic challenge that comes across while crowdsourcing is the resources that must be available with the community. Though, in the past decade, percentage of mobile phones available across the world has increased, but to attain all the valuable information from the community, from the places, where the other public resources are not available is a big challenge with unavailability of the resources other than mobile. When we talk about the information crowdsourced in case of quality of air pollution or about tap water then various other technologies are required other than a Mobile. Like the idea of ParkSense depend highly on wifi beacons and fitting every parking spot with this technology is a tedious task.
* **Privacy issue:[39]** This is also a major concern for the community. With growing technology of AI and ML, communities that are participated into online surveys and contests have a very big chance of being attacked by malicious user on basis of the data about their liking and disliking. This issue has been brought out by various users.
* **Data Authenticity and redundancy:[39]** It was another a very common challenge faced during the studies. Users when interacted with the real-time apps and platforms to update the alerts, some did just for the sake of it while some do it without the proper knowledge of using the devices. In case of Air quality index testing, readings taken when holding the device in hand is more accurate than the readings when the device is in the pocket[3]. Multiple users in the same area can report same incident that leads to redundancy of data and loss of resources in transferring it over the internet and storage.
* **Sparsity of data[40]:** When community crowdsourced a lot of data there are chances when data for some entities is not received from anyone and this is widely observed in crowdsourcing data. Some domains that are generally wished to be explored are left aloof because of this issue. The inconsistency in receiving the data is also the hurdle in the study which made the data more complex
* **Unwillingness of the crowd to participate[42]:** People are generally organised into groups, communities and urban groups. These are generally those who have either common interest in that topic or they share a common problem regarding that study. To get regularly updated with correct data is very unreliable in some of the cases too. User security and privacy issue also demotivates them to continue the support.
* **Robustness and stability[22]:** In many studies, after focusing on all the challenges, comes a most testing part i.e the robustness and stability of their system. How will the system behave after taking the platform live is one of the challenge that is not kept in mind initially and when everything starts to take off the performance expected from the system starts to lag behind.
* **Testing phase[22][23] :** Every research focuses on the challenges and steps that they should take but nobody discussed about the complexities that will come ones the system goes online. Testing the whole concept with a huge amount of community sending information and data is what has been overlooked generally.

These were the general problems that came across our research. The motive of our research was not only to note them but also to address them. Here, we talked about the two of the most common challenges that have been going rounds all over the internet as well in research world. So, in our thesis we will address the problem:

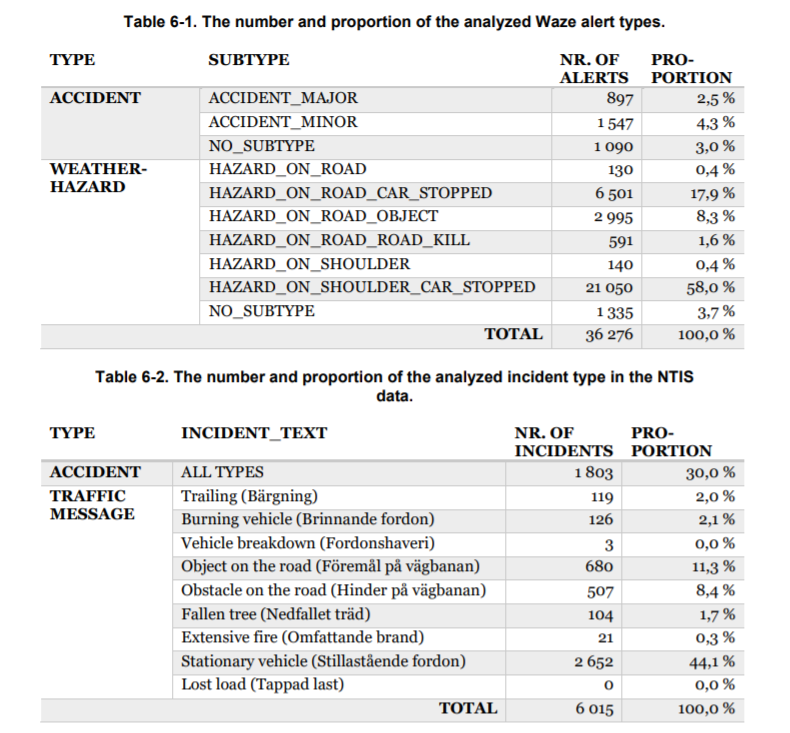
**Research questions:**

1. Redundancy and Authenticity of the data:
   1. How to deal with incorrect, fake and unreliable data?
   2. Which applications have been used to overcome the problem?
   3. How much data can be consistent?
   4. How much is the data effective after elimination?
2. Unwillingness of the crowd to take part due to various reasons.
   1. What are the challenges that may stop participants to participate?
   2. What are the types of incentives?
   3. How much beneficial is the outcome if the incentives are intrinsic or extrinsic?
   4. What is the direct motivation for the participants without incentives?

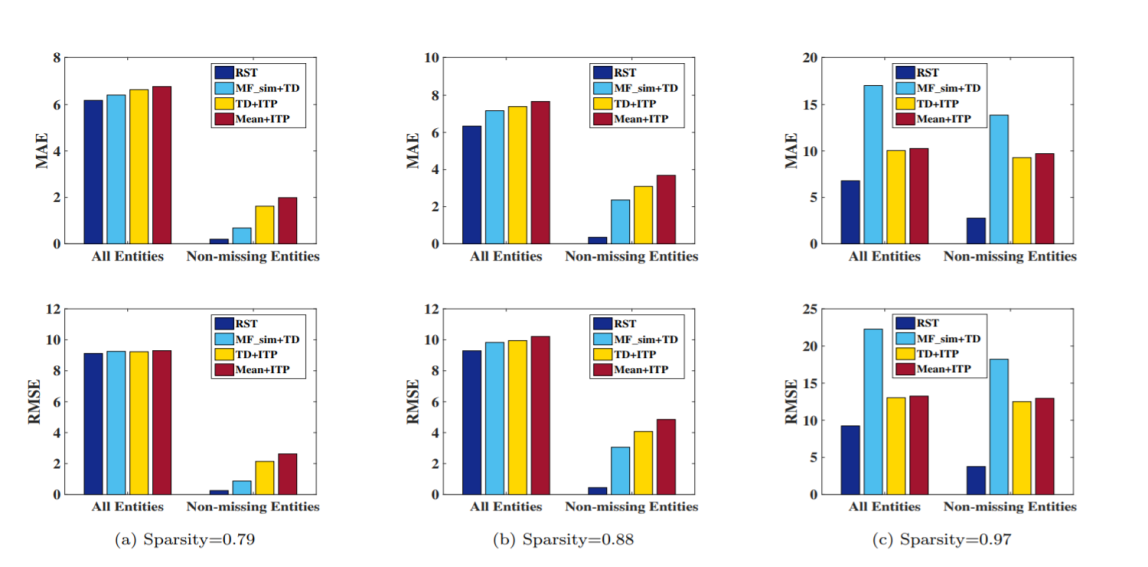
Result

The basic questions about the challenged in data crowdsourcing that we explored are well explained in this part of the thesis. As we are addressing data redundancy and crowd participation we came through various studies and found these some answers that we think are optimal.

1. Data redundancy:
   * + 1. Waze Platform [43]:
          1. It is a navigation application that used to crowdsourcing concept and talked about the challenges in which one of the challenges is data redundancy and they have tried to optimize the data authenticity. They have tried matching their data with the departmental data and from other resources to minimize the redundancy but the result came after matching their data with other resources were 4.5 times higher than the actual registered incidents with the Department. They used a techique discussed in [44].
          2. It was also found that, 27.5% of the incidents could have been received early by using the crowdsourced application. But they also got redundant data. So, to segregate redundant and reliable data they assess the reliability of their users by ranking them. As the user data, who is reporting the incident do not have personal information to assess its reliability. So, they decided the authenticity on the basis of the number of confirm alerts reported by other regular users. The maximum number of alerts on a particular incident increases the value of that alert and places that in reliable alerts list. Then those particular alerts were aggregated.



* + - * 1. A logistic regression method was used to find the effects of the independent variables on the probability that a Waze alert matches to an incident which has been reported with the department. The analyzed independent variables were the type of the alert, type of road where the incident happened, the part of the day when the incident happened, the rank of the reporter, the number of related alerts, the number of confirmations that the alert has received. After that the alerts and confirmations reported within 5 minutes were analyzed and thus some redundancy was controlled.
      1. Crowd sensing application.[45]:
         1. This study was done in tackling of redundancy in mobile sensing Technology and they also faced this issue[46]-[49]. In this they worked upon a framework to estimate the true values from each user. As, multiple users can observe and send the observations that can be redundant as well as conflicting with each other because of unreliable nature of the sensors when carried by humans. At the peak hours of the day users at some popular locations reports redundant data. They proposed an algorithm in their framework that aggregates both user-contributed and inferred observations to discover the actual observations.
         2. There is an approach to conduct voting or averaging the values that are claimed by majority of the users. But, the drawback of doing this was it overlooked the reliability of the users. The other reason for the conflicting values was the hardware and the technique that user applied while sending the observations as values could be different when sensor is in the open v/s when it is in the pocket.
         3. They also tried the approach of user reliability by truth discovery as aggregating user on the basis of observations. But this was also lacking at some points when users at a particular place was not substantial to aggregate the observations. So they Finally develop Redundancy Tackling (RT) [45] framework, to infer the true values of users from redundant and sparse data in crowd sensing applications. This design an optimized method that extracts key information not only from user observations but also from the similarities between them and recover a complete user observation matrix which consists of matrix factorization and regularization terms.
         4. An effective solution is developed to solve the proposed optimization problem in an iterative way. The convergence property of the proposed solution is proved, and effective techniques are presented to further reduce the time complexity.



* + - 1. We also came around 2 more past studies [50] that were a not recent but they came into the light just because of the least amount of budget requirements to execute these surveys. They have also faced the same challenges of data redundancy and inconsistency.
         1. A study of Movi [51] , a view selection module was developed to select the most high quality videos, complementing the multidimensional sensing to obtain the “best view” of the recorded event. Like for the case of readings of acetometers for selecting stable images.
         2. In this model a complex issue was derived of inconsistent multimodal sensory data as both the video or audio clips could be used to but the inference results might be different, so, further studies were required in that domain to address the inconsistency that might caused by the multimodal data contributed by the crowds.

So, through these studies we can confer that redundancy is one of the most common challenges that are faced while data crowdsourcing, and no matter what technique is applied there is always some areas that lags behind. But keeping the different approaches in mind we conclude that data redundancy was controlled by comparing and merging the data with other sources, and applying the regression method with various variables, there was still a scope of improvement in the final readings but the one used with RT framework with truth discovery techniques was better.

1. Crowd participation:
   * + 1. During the research on crowd participation and motivation, a thesis suggest various aspects of incentivisation[50]. They found some studies that focused on the development models of incentivization to facilitate data sharing. It was found that incentives can be distinguished into 2 types. Intrinsic and financial incentives.
          1. Intrinsic incentives: They are where volunteers are interested to help when they think about the task as interest and important, or enjoyable, or where the participators are attracted of having a chance of receiving public recognition. These incentives worked well in environment monitoring cases.
          2. Financial incentives: This is the easiest way to motivate user participation in almost all types of cases. Financial rewards to their participation could be money, virtual cash, redeemable coupons. But the drawback that was observed was the human nature of deceiving where the money is involved, to gain the most. So more studies were referred to on how to rewards the most truthful contributions.
          3. So, several game methods was preferred but all were nipped in the bud due to complex implementation of them in a fully distributed time evolving system.
       2. Another study was found which revolves around the “How to motivate crowd participation” in online crowdsourcing platforms. As small and big businesses requires online business, crowdsourcing had been their one stop solution to concentrate on online customers. The main risk was extinction of active members.
2. Through this paper [52] they try to understand what motivates the people to participate in online crowdsourcing platforms.
3. It was noted crowd had two types of motivation: intrinsic and extrinsic. Where they are motivated by their own interests or hobbies and the other were they expect something in return.
4. According to them, In crowdsourcing platforms, the influential extrinsic motivational factors are reputation, status, peer pressure, fame, community identification and fun. They have also found that financial rewards are harmful for idea creation whereas [53] it was also discovered in their research that monetary rewards are essential.
   * + 1. This study was related to water contamination levels in the city[54], where all of private drinking tap (home, kiosks, community places) were monitored by water samples by crowd. The project has also talked about the citizens motivation to take part in the initiative to provide samples daily. But only 7% of the users had completed the survey till the end. They motivated the citizens to participate in the project due to concerns about their health or their families. Participants though has reported that they were not motivated by the cash rewards.

They organised various contest in an effective way to attract a broader audience and to generate more desirable solution. They used a pre and post survey design to describe patterns and identify the outcomes from participation in the project. It was to ask questions on how they found out about the project and how they keep them motivated etc. They also built from the Developing, Validating, and Implementing Situated Evaluation Instruments (DEVISE) Framework for motivations to participate in crowdsourcing.

Two contest schemes were also designed to test the role of incentives i.e Go-Getter” and “Ambassador,” each with $200 cash rewards. each rewards the participant who collects the most sample.

A total of 136 pre-survey responses and 42 post-survey responses had more than 50% questions answered, and hence were included in their survey analyses. When they were asked if they would be willing to participate in another project like this if given the opportunity, 77% of participants said “Yes” and 19% s said “Maybe.”

After the experiment and the 2 contest “GoGetter” and “Ambassador”, the participants who completed the whole survey told they came to know about the survey through social media or word-of-mouth.

This shows the limitations of cash incentives and social media recruitment in a drinking water monitoring/testing project. Meanwhile, 53% of participants told that about the project when they saw a kiosk in person. This implies that persons are more likely to participate if the project materials are directly accessible to them and providing further evidence that health and identity related factors were key factors that motivated participation in the contest.

Around 67% of participants either disagreed or strongly disagreed that they were motivated to participate by the cash prize, and 21% were neutral. Additionally, when asked which contest the participant hoped to win, 82% of participants responded “Neither” in the pre-survey. For those who participated and also responded to the postsurvey, 92%) indicated they were not participating in either of the contest option.

A screenshot of a cell phone

Description automatically generated

Hence, It is clear from this last study, that this is the best optimum solution, we have found because of the less extrinsic factor that comes to play in environmental monitoring surveys.

Conclusion

After reading all the studies related to data Crowdsourcing and looking into the common challenges faced by those studies, it has been concluded that the redundancy, inconsistency and reliability of data is their major challenge for any environment monitoring or survey-based projects.

Either it is the Waze platform [43][44] that uses user reliability approach combined with logistic regression method to analyze the independent variable so to aggregate the redundant data or it can be the second study of all sensing equipment’s in general which built a Redundancy tackling framework to infer the true values from user observations and from other sources. They also had the similar approach of truth discovery like Waze on the basis of observations, but their matrix factorization has helped in a better to address those redundant data and aggregate them keeping in view the observation nature of every participant and not putting everyone’s work under same incident category.

The technique used by Waze in traffic control monitoring [44]was useful upto a limit but looking at their readings in the result section, it can be inferred that the technique they used in that project does not gave the best results as the no of data alerts was way more than departmental data when it was compared, after they applied their approach.

But the other one was quite fruitful with this technique in their work. This has not only provided better results for redundancy as well as solved sparsity challenges that were not addressed by us but this result has also considered every user’s reliability variable.

The other challenge that most surveys and projects came around is how to encourage participation of the user till the end of the project. Various projects worked on the crowd motivation techniques and advertisements while others look into the scope of incentive measures that could be possible for keeping the survey alive.

The first study[50] described the role of personal motivation that can keep the survey running or the contest methods to encourage their participations, which was explained by the third study that was done with water contamination, that most of the users that completed the survey has never take part in the contest.

The second study[52] [53] has worked on the users crowdsourcing in online platforms for businesses. They had the same result like the first one where it was noticed that either the user was motivated only because of interest or hobbies. It was also stated that incentive approach is like virtue and vue.

But the third study[54] was what is worth considering, It worked on all aspects of using crowd motivations, organising crowd contest, giving incentives etc. The surveys done by them on users after the project completion put light upon the participant preception that can help the other projects like this.

So, if crowd participation has to be increased in environment monitoring projects, then the strategy used by this third study is the best one. It explains that there should be mixed approach with more focus on crowd motivation because of the health hazards upon their family members and themselves. The other possible way to motivate them was incentive contests and games that only impacts for a smaller portion of time in the whole project but this was difficult to apply in bigger projects that evolves with time. So this is an area that was not explored and in future it can be worked upon to get the real picture.

**Scope and future work:**

Crowdsourcing is developing since the its inception. It can’t be achieve fully in a single project. It can help to solve lack of information and data regarding a problem. Though the studies that were successful in finding a way to solve their problems but there is always some scope of improvements with other various aspects that they overlooked in their study. In future, for such type of crowdsourced projects their method of technique can be followed while exploring some new areas of work together. These methods was only prescribed for specific projects needs but for the vast area of business and lack of information for their optimisation, their more study is needed to fully exploit its potential. In many years to come we can look forward to AI (Artificial intelligence) and ML (machine learning) to help this domain to achieve more.

References

[1] S1.pulso.cl, 2020. [Online]. Available: http://s1.pulso.cl/wp-content/uploads/2016/02/2258581.pdf.

[2] "The global footprint of mobile communications: The ecological and economic perspective - IEEE Journals & Magazine", Ieeexplore.ieee.org, 2011. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs\_all.jsp?arnumber=5978416.

[3] "Crowdsourcing with Smartphones - IEEE Journals & Magazine", Ieeexplore.ieee.org, 2012. [Online]. Available: https://ieeexplore.ieee.org/document/6216342/.

[4] "Crowd Computing: A Literature Review and Definition", Dl.acm.org, 2013. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/2513456.2513470.

[[5] P. Neis, D. Zielstra and A. Zipf, "The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011", 2012.

[6] "Vehicular Social Networks: Enabling Smart Mobility", Ieeexplore.ieee.org, 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7926916.

[7] D. Stojanovic, B. Predic and N. Stojanovic, "Mobile crowd sensing for smart urban mobility", 2016. [Online]. Available: https://www.ubiquitypress.com/site/books/10.5334/bax/download/942/#page=382.

[8] "Towards Smart City | Proceedings of the 18th International Conference on Distributed Computing and Networking", Dl.acm.org, 2017. [Online]. Available: https://dl.acm.org/doi/10.1145/3007748.3018286.

[9] "Transafe: a crowdsourced mobile platform for crime and safety perception management - IEEE Conference Publication", Ieeexplore.ieee.org, 2011. [Online]. Available: https://ieeexplore.ieee.org/document/7160600/.

[10] "Mobile Phones as Seismologic Sensors: Automating Data Extraction for the iShake System - IEEE Journals & Magazine", Ieeexplore.ieee.org, 2013. [Online]. Available: https://ieeexplore.ieee.org/document/6470715/.

[11] X. Chen, E. Santos-Neto and M. Ripeanu, "Crowd-Based Smart Parking: A Case Study for Mobile

Crowdsourcing", 2012. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-36660-4\_2.

[12] "Harnessing the crowds for smart city sensing | Proceedings of the 1st international workshop on Multimodal crowd sensing", Dl.acm.org, 2012. [Online]. Available: https://dl.acm.org/doi/10.1145/2390034.2390043.

[13] "Large-Scale Mobile Sensing Enabled Internet-of-Things Testbed for Smart City Services - Jorge Lanza, Luis Sánchez, Luis Muñoz, José Antonio Galache, Pablo Sotres, Juan R. Santana, Verónica Gutiérrez, 2015", SAGE Journals, 2015. [Online]. Available: https://journals.sagepub.com/doi/full/10.1155/2015/785061.

[14] T. Franke, P. Lukowicz and U. Blanke, "Smart crowds in smart cities: real life, city scale deployments of a smartphone based participatory crowd management platform", 2015. [Online]. Available: https://link.springer.com/article/10.1186/s13174-015-0040-6.

[15] "The relationship between road safety and congestion on motorways", Swov.nl, 2012. [Online]. Available: https://www.swov.nl/sites/default/files/publicaties/rapport/r-2010-12.pdf.

[16] "A Web of Things based eco-system for urban computing - towards smarter cities - IEEE Conference Publication", 2017. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7998277.

[17] "George Hotz", En.wikipedia.org. [Online]. Available: https://en.wikipedia.org/wiki/George\_Hotz.

[18] "comma.ai – Introducing openpilot", comma.ai – Introducing openpilot. [Online]. Available: https://comma.ai/.

[19] N. Maisonneuve, M. Stevens, M. Niessen and L. Steels, "NoiseTube: Measuring and mapping noise pollution with mobile phones", 2009. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-540-88351-7\_16.

[20] Dl.acm.org, 2013. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/2500423.2500438.

[21] "BikeNet: A mobile sensing system for cyclist experience mapping: ACM Transactions on Sensor Networks: Vol 6, No 1", Dl.acm.org, 2010. [Online]. Available: https://dl.acm.org/doi/10.1145/1653760.1653766.

[22] "Large-scale complex IT systems | Communications of the ACM", Dl.acm.org, 2012. [Online]. Available: https://dl.acm.org/doi/10.1145/2209249.2209268.

[23] G. Cugola and A. Margara, "Deployment strategies for distributed complex event processing", 2013. [Online]. Available: https://link.springer.com/article/10.1007/s00607-012-0217-9/.

[24] J. Marynowski, A. Santin and A. Pimentel, "Method for testing the fault tolerance of MapReduce frameworks", 2017. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1389128615001358.

[25] "What is Crowdsourcing? | Crowdsourcing Week", Crowdsourcing Week, 2020. [Online]. Available: <https://crowdsourcingweek.com//what-is-crowdsourcing>.

[26] "History of Crowdsourcing & Why You're Just Hearing About It", CrowdSource, 2013. [Online]. Available: <https://www.crowdsource.com/blog/2013/08/the-long-history-of-crowdsourcing-and-why-youre-just-now-hearing-about-it/>.

[27] A. Fournier, "How to Boost Company Innovation with Crowdsourcing | Braineet", Braineet, 2019. [Online]. Available: <https://www.braineet.com/blog/crowdsourcing-innovation/>.

[28] "Open Innovation", Unilever global company website. [Online]. Available: <https://www.unilever.com/about/innovation/open-innovation/>.

[29] Fructose-1,6-Bisphosphatase 2 Inhibits Sarcoma Progression by Restraining Mitochondrial Biogenesis", 2016. [Online]. Available: 1. https://www.sciencedirect.com/science/article/abs/pii/S0043135419309832#abs0015.

# [30] Crowdsourcing: A Review and Suggestions for Future Research, 2017. [Online]. Available: <https://www.researchgate.net/publication/309642734_Crowdsourcing_A_Review_and_Suggestions_for_Future_Research>.

[31]"Crowdsourced traffic information in traffic management", Kth.diva-portal.org, 2018. [Online]. Available: <http://kth.diva-portal.org/smash/get/diva2:1266883/FULLTEXT01.pdf>.

[32]"Waze Communities — meet the people behind the technology", Waze.com. [Online]. Available: <https://www.waze.com/communities>.

[33] "How does Waze work? - Waze Help", Support.google.com. [Online]. Available: <https://support.google.com/waze/answer/6078702?hl=en&ref_topic=9022747>

[34]"Waze Map Editor", Waze.com. [Online]. Available: <https://www.waze.com/en-US/editor?lat=59.33205&lon=18.05191&zoom=2>

[35] D. Rudmark, "Att öppna datakällor för tredjepartsutveckling - rekommendationer för kollektivtrafikbranschen", DIVA, 2018. [Online]. Available: http://www.diva-portal.org/smash/record.jsf?pid=diva2:1131851.

[36]"Crowdsourced traffic information in traffic management", Pdfs.semanticscholar.org, 2018. [Online]. Available: <https://pdfs.semanticscholar.org/8a02/5af73998a7fd7f662abe484828a47c85968c.pdf>.

[37] "Incident Traffics Management Response", Scholarworks.waldenu.edu, 2017. [Online]. Available: <https://scholarworks.waldenu.edu/cgi/viewcontent.cgi?article=4898&context=dissertations>.

[38] "Wazestats - Waze Statistics", Wazestats.com. [Online]. Available: <http://wazestats.com/active.php>.

[39] https://link-springer-com/article/10.1007/s10606-020-09374-0[39] - Google Search", Google.com, 2020. [Online]. Available: <https://www.google.com/search?client=safari&rls=en&q=1.+https://link-springer-com/article/10.1007/s10606-020-09374-0%5B39%5D&ie=UTF-8&oe=UTF-8>.

[40] Cse.buffalo.edu, 2016. [Online]. Available: <https://cse.buffalo.edu/~lusu/papers/SenSys2016.pdf>.

[41] P. Huangyang et al., "Fructose-1,6-Bisphosphatase 2 Inhibits Sarcoma Progression by Restraining Mitochondrial Biogenesis", 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0043135419309832#abs0015>.

[42] "Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm: ACM Computing Surveys: Vol 48, No 1", Dl.acm.org, 2015. [Online]. Available: https://dl.acm.org/doi/10.1145/2794400.

[43] Kth.diva-portal.org, 2017. [Online]. Available: http://kth.diva-portal.org/smash/get/diva2:1266883/FULLTEXT01.pdf.

[44]"Santos, S., Davis Jr., C. A., & Smarzaro, R. (2017). Analyzing Traffic Accidents based on the Integration of Official and Crowdsourced Data. Journal of Information and Data Management, 67-82.[44] - Google Search", Google.com, 2017. [Online]. Available: https://www.google.com/search?client=safari&rls=en&q=1.+Santos,+S.,+Davis+Jr.,+C.+A.,+%26+Smarzaro,+R.+(2017).+Analyzing+Traffic+Accidents+based+on+the+Integration+of+Official+and+Crowdsourced+Data.+Journal+of+Information+and+Data+Management,+67-82.%5B44%5D&ie=UTF-8&oe=UTF-8.

[45] Cse.buffalo.edu, 2016. [Online]. Available: https://cse.buffalo.edu/~lusu/papers/SenSys2016.pdf.

[46] "A mobile sensing system for cyclist experience mapping: ACM Transactions on Sensor Networks: Vol 6, No 1", Dl.acm.org, 2009. [Online]. Available: <https://dl.acm.org/doi/10.1145/1653760.1653766>.

[47] "Scaling Crowdsourcing with Mobile Workforce: A Case Study with Belgian Postal Service: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies: Vol 3, No 2", Dl.acm.org, 2010. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3328906.

[48] [11]"EasyTracker: Automatic Transit Tracking, Mapping, and Arrival Time Prediction Using Smartphone’s”,", Citeseer.ist.psu.edu, 2011. [Online]. Available: <http://citeseer.ist.psu.edu/showciting?cid=13764112>.

[49] [12]"AirCloud | Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems", Dl.acm.org, 2014. [Online]. Available: <https://dl.acm.org/doi/10.1145/2668332.2668346>.

[50] "Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm: ACM Computing Surveys: Vol 48, No 1", Dl.acm.org, 2017. [Online]. Available: <https://dl.acm.org/doi/10.1145/2794400>.

[51] "Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm: ACM Computing Surveys: Vol 48, No 1", Dl.acm.org, 2010. [Online]. Available: https://dl.acm.org/doi/10.1145/2794400.

[52] "IEEE Xplore", Ieeexplore.ieee.org. [Online]. Available: https://ieeexplore-ieee-org.ezproxy-f.deakin.edu.au/abstract/document/6236409.

[53] "MOTIVATIONAL FACTORS AFFECTING PARTICIPATION AND CONTRIBUTION OF MEMBERS IN TWO DIFFERENT SWISS INNOVATION COMMUNITIES | International Journal of Innovation Management", Worldscientific.com. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/S1363919611003477.

[54] “P. Huangyang et al., "Fructose-1,6-Bisphosphatase 2 Inhibits Sarcoma Progression by Restraining Mitochondrial Biogenesis". [Online]. Available: https://www.snfge.org/content/httpwwwsciencedirectcomsciencearticlepiis0016508516300543.