

## **Studying internet penetration in to Rural India by interpolating crowd sourced cellular data points of different telecom operators.**

(Mitesh Patel, Samta Sharma, Kapil Chaudhary)

### **Abstract:**

Rural India is the backbone of our country. Any development would not become much effective unless contained from rural India. With our wish to become digital country it is the access to internet which counts most. Various surveys show that 75% of new Internet users in India will come from rural India. Mobile internet penetration is estimated to grow 56% YoY. It has been forecasted that 71% of the mobile internet users in India will hail from urban areas and rest from Rural area. Buoyed by Internet penetration in rural areas, the number of web users in India will see a two-fold rise at 730 million by 2020. We could capture past trails and predict future growth of internet into rural areas using Geospatial data. In this paper, we are presenting one such study. We utilised Geospatial technology to study the ground level facts. We have collected massive crowd sourced point data of various radio signal measurements from Opencellid.org for past few years. In order to create time series data, we used various interpolation techniques available in MapInfo Pro Advanced to process this massive point data and generate the heat maps. These heatmaps are then stored as temporal MRR (Multi Resolution Raster) image to run further analysis tools. This enables us to visualise and analyse past and current trends to validate how rural areas are progressing with internet adoption. We extracted key information in the form of geospatial images like which generation (2G,3G,4G) of mobile internet is consumed most, differential maps between different time points to find adoption levels, availability of service by different Telco, and clusters of network availability.

### **Back Ground:**

As we are progressing towards open internet, we need to validate our readiness of Rural India to absorb, understand and use the Internet Technology. We need to know the velocity at which India and especially Rural India is absorbing this technology. As an attempt, we looked for some data to analyse upon. We needed aggregated data about accessibility to mobile internet by Indian citizen. This data is treasure for the Telecom operators to put it in to public domain. We used crowd sourced telecom measurement data from OpenCellId for this purpose. We have used Mapinfo Pro Advance's spatial analysis capabilities to analyse these measurements and find trends and patterns in it Like how fast internet access is spreading, how fast is the switch from 2G to 3G to LTE by analysing the historical data.

### **OpenCellID:**

As per the OpenCellID.org<sup>1</sup> "OpenCellID is the world's largest collaborative community project that collects GPS positions of cell towers, used free of charge, for a multitude of commercial and private purposes." It is a crowd sourced data collected by contributors through mobile apps installed on their devices. Basic aim is to locate every single cell tower in the world. It also collects other useful information like, Radio Signal strength, Type of network, Network Provider, etc. We have used all of this information to analyse mobile internet Growth

in India. OpenCellID mainly stores this information in CSV format created with the help of Mozilla Location Service<sup>2</sup>.

### **Approach:**

Whenever we start any geospatial analysis, we get a crucial question, whether to analyse the available geospatial information as in to Vector domain or in to Raster domain. Many times we do not bother and try to use data as it is for the solution i.e. if we have vector data at the hand we mostly try to solve the problem first using vector analysis. Actually, it should not be the user's problem to choose the right domain.

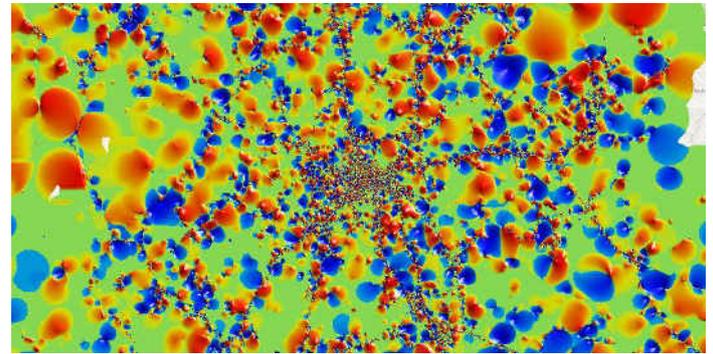
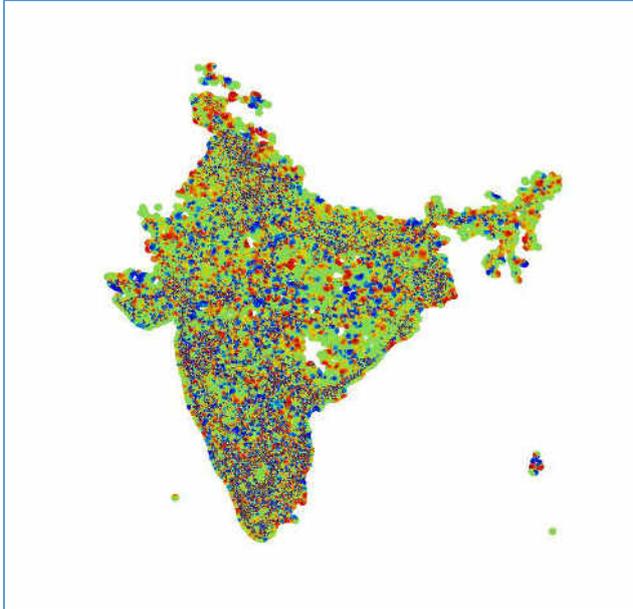
For this particular scenario, we will use fusion of Vector and Raster analysis domain. Our source data is CSV files with location in the form of Lat/Long in it, which could be considered as a vector data source. We transformed this vector source data in to a continuous raster surface using MapInfo Pro Advanced<sup>TM</sup> Interpolation Tools. This will give us continuous Raster which could then be used as a source for further analysis. E.g. Using Calculator Tool, we could find out regions where network coverage was not present before some specific date, these could reveal the regions which were recently connected to the Mobile internet. Similarly, we can identify the area's where network is upgraded e.g. from 2G to 3G. We can find rate at which mobile internet is spreading.

Finally, we could query these rasters at specific point to create vector output. Like we can do point inspection with village points to identify if mobile internet is available in those villages. Or we could create regions out of this raster using Vectorise tool to extract all the regions with some particular type of network 3G or LTE. This brings us back in to the Vector analysis domain. Just see how we transitioned from initial vector inputs to Raster domain for analysis and back in to Vector domain to query, present and showcase the outcome.

In upcoming sections, we will be discussing some of these analysis and outcomes.

### **Aggregated Network Coverage Map:**

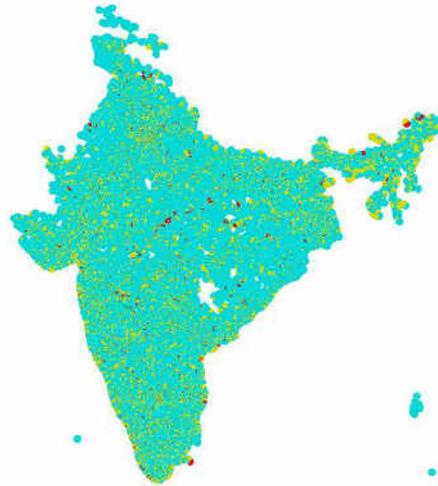
In this analysis, we have prepared aggregated network coverage map as a Raster surface. We have used measurement locations from the recent measurements CSV file from OpenCellID data as a source data. We used Lat and Lon attributes out of it to create a Point cloud. We then interpolated this point layer to create a continuous raster using Create Raster tool in Mapinfo Pro Advanced. We have used Natural Neighbour interpolation algorithm to interpolate this data at 500M cell resolution. We used Signal Strength attribute for interpolation giving us coverage map as a continuous raster aka (Heat Map) shown in Fig a. Reds are hot (good signal strength) and blues are warm (Poor signal strength).



**Fig a. Aggregated Network Strength Map (Left)  
Rural Vs Urban cell range (Right)**

### **Aggregated Network Type Map:**

For this analysis, we used Network type (2G,3G, LTE) information from OpenCellID measurement file. Again Lat, Lon attributes were used to create Point cloud. We used radio attribute, which basically is of string type attribute representing type of Radio (2G,3G, LTE). This attribute then is transformed to numeric field type with some type number assigned to it using SQL query tool. This numeric type number is then used to create raster surface using Nearest Neighbour algorithm. Which actually assign values to raster cells with value of nearest data point's value. This is in contrast to heat map where data point's values are actually interpolated and then assigned to a cell. This is the reason we have used Nearest Neighbour algorithm this time, as we wanted raster cells to have assigned actual data point value. Other choices of algorithm in this scenario could be stamping algorithm. The output raster is shown in Fig b.



**Fig b. Aggregated Network Type (2G = Light Green, 3G = Lemon Green, 4G = Red)**

### **Network Operator Map:**

Another interesting aspect to analyse is network operator. Also, we could analyse the patterns in Rural VS Urban and Private Vs State Owned Operator, who is catering to which consumers the most. We could also see which operator has a strong hold in a particular region. We could also estimate coverage area under individual network operator. OpenCellID data also has Mobile Network Code (MNC) details, which can be used to identify the operator. Again, these MNC codes were aggregated using Mapinfo Pro's SQL query, as each operator will usually have tens of MNC codes. We then created raster surface using nearest neighbour algorithm, as we need exact data point values to be assigned to each raster cell (Pixel) in this case as well.

Once we have raster ready we could simply classify this raster and use the statistics of classified raster to measure foot prints of individual Network Operator. Alternatively, we could create vector regions using vectorise tool, which creates vector regions for individual network operators.

Fig c. shows output raster showing network operator map.

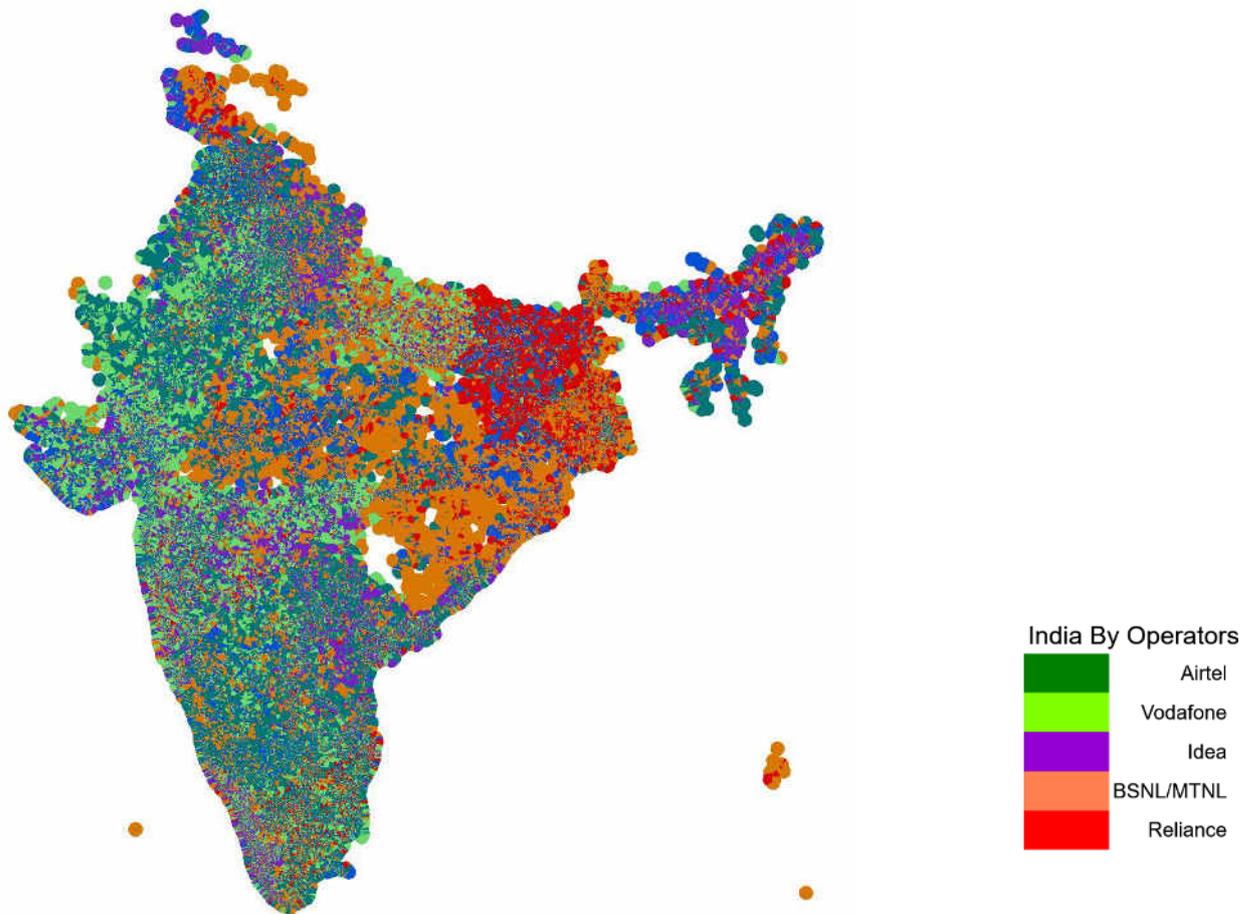


Fig C. Converge by Network Operator (Left), Legends (Right)

### Rate of Mobile internet penetration:

There are various ways we can analyse OpenCellID data to see how mobile internet is spreading. We could even compare how particular radio network type is expanding or shrinking, by comparing historical data between different time points.

In our analysis, we have compared accumulated measurements between 2015 and 2016 to find the patterns of such changes.

We used calculator tool to extract area under particular network type, 3G or 4G between year 2015 and 2016. Four rasters were created covering each network type for each year e.g. 3G coverage in year 2015. Once rasters are crated, we can get aggregated area from raster statistics. Fig. b shows bar graphs created from these statistics to see the comparison and compare the growth rate.

As it looks obvious from fig d. that in 2016 people used more of 3G network than in 2015. Growth rate could be estimated by line connecting both the bars. What is more encouraging is 4G adoption is much steeper than 3G adoption between year 2015 and 2016 asserting the various reports we mentioned earlier.

Fig. e shows base raster created for each year for the aggregated network type covering all the operators.

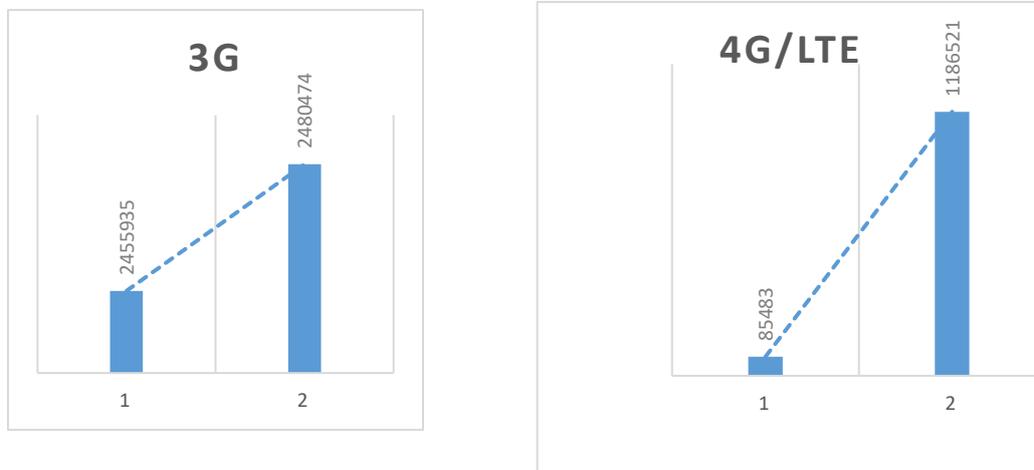


Fig d. difference in coverage area of particular type (3G,4G) between 2015 and 2016

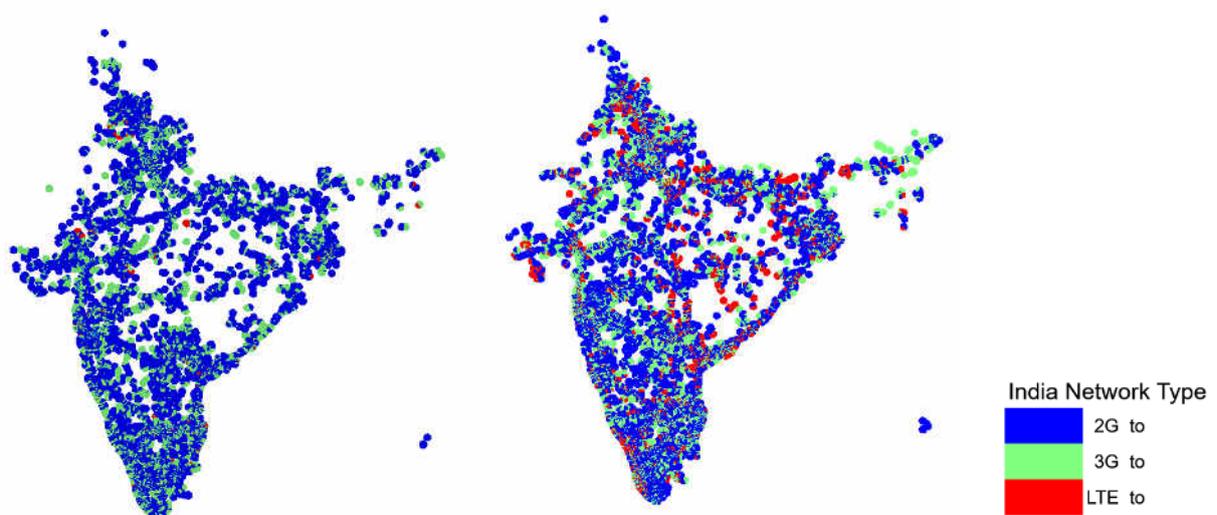


Fig e. Network Type 2015 (Left), Network Type 2016 (Centre), Legends (Right)

#### Availability of mobile network at specific points:

Once we have network heat maps ready with us, we could use it for further analysis. Like, for set of input points, which could be representing some village points or any other probing points, we could run point inspection tool to know whether we have network coverage available under that point or not. We could also query type of network coverage available under the points.

### **Conclusion:**

We analysed mainly three type of measurements submitted by individual volunteers/users of OpenCellID.org namely Network Type, Signal Strength and cell range. We also looked at network operator (MNC) for comparison. Different CSV files downloaded from OpenCellID were converted to point cloud and then continuous raster surface was generated using interpolation to have continuous data and fill the holes in data reasonably.

Through this analysis, it came out that Internet penetration is continuously increasing in to both urban and rural areas in India. Newer technologies like 4G/LTE is well received by end users than others (2G, 3G). This gives a confidence that Indian citizen are ready and more eager to consume latest internet technologies if serviced at right value.

### **References:**

1. OpenCellID: [http://wiki.opencellid.org/wiki/What\\_is\\_OpenCellID](http://wiki.opencellid.org/wiki/What_is_OpenCellID)
2. MozillaLocationService: <https://location.services.mozilla.com/>
3. Mapinfo Pro Advanced: <https://www.youtube.com/user/mapinfo>  
<http://www.pitneybowes.com/us/location-intelligence/geographic-information-systems/mapinfo-pro.html>