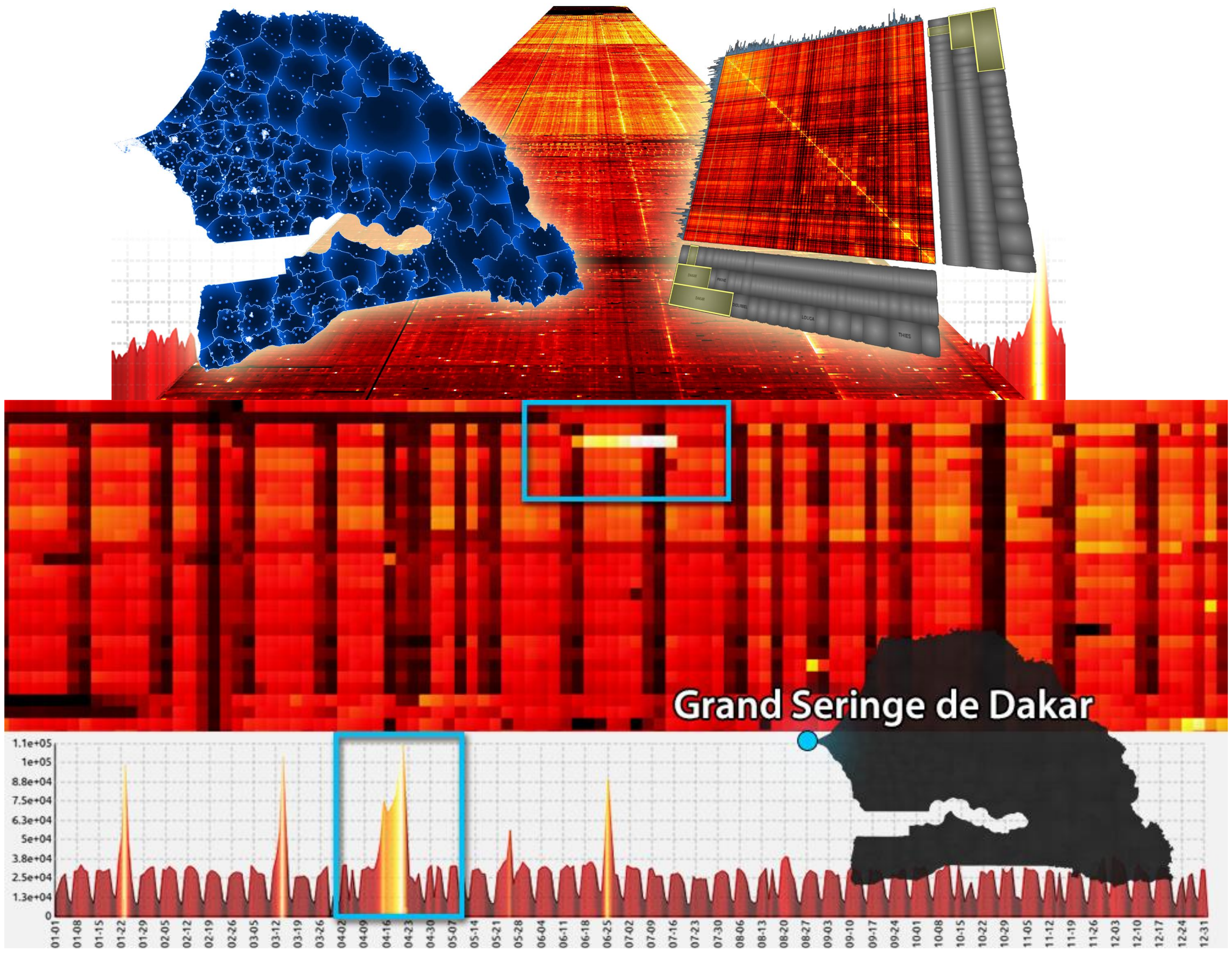
Data for development reloaded: visual matrix techniques for the exploration and analysis of massive mobile phone data

O05

Data for Development Reloaded: Visual Matrix Techniques for the Exploration and Analysis of Massive Mobile Phone Data

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



Stef van den Elzen[1,2], Martijn van Dortmont[1,2], Jorik Blaas[2], Danny Holten[2], Willem van Hage[2], Jan-Kees Buenen[2], Jarke J. van Wijk[1], Robert Spousta[3], Simone Sala[3], Steve Chan[3], Alison Kuzmickas[3]

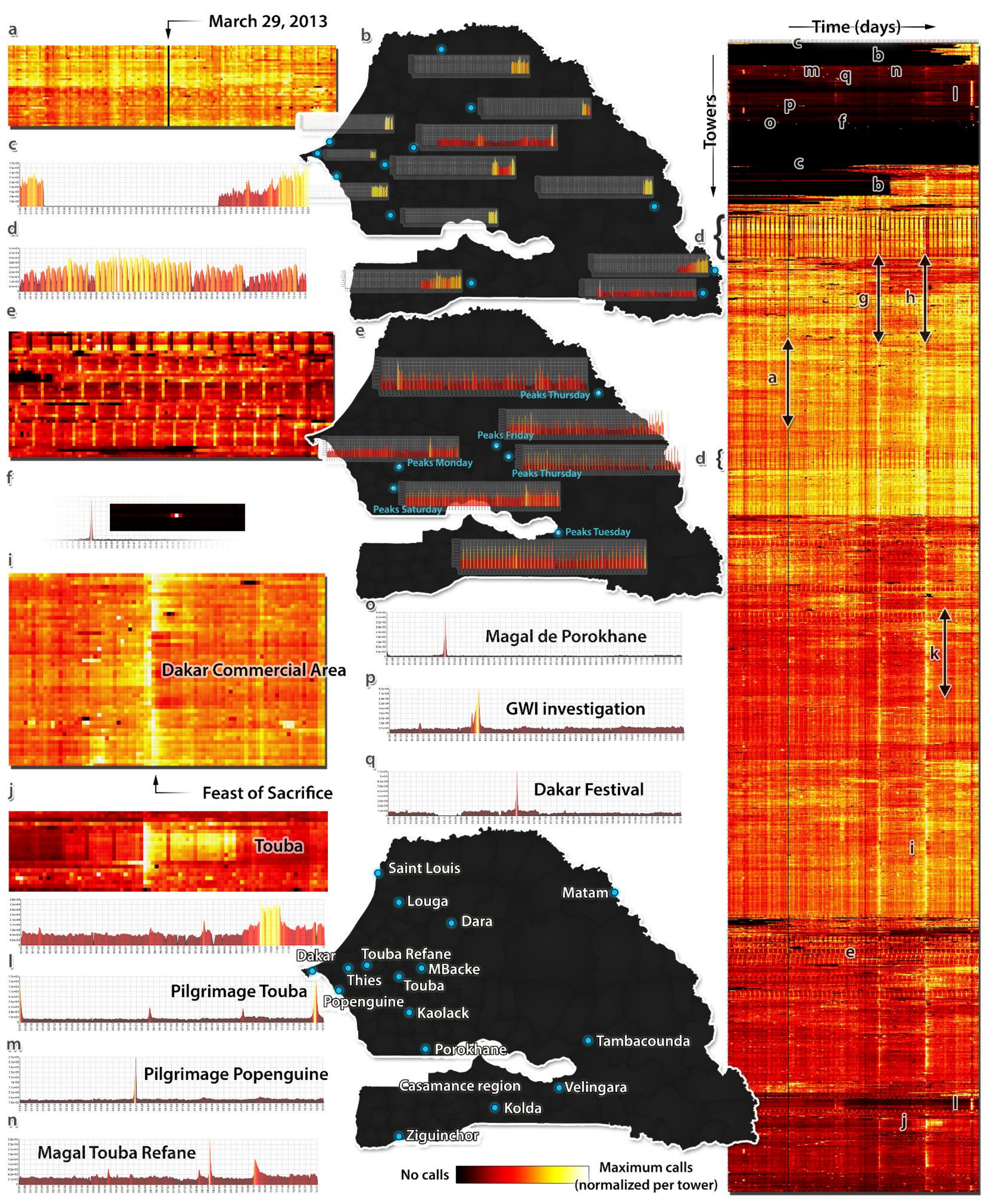
(1) Eindhoven University of Technology (2) SynerScope BV (3) Sensemaking Fellowship

Project Summary:

We present visual analytics techniques for the exploration and analysis of massive mobile phone data. We use a multiple coordinated view approach with a scalable and flexible visual matrix as central element to our solution. Users are enabled to identify both temporal and structural patterns such as normal behavior, outliers, anomalies, periodicity, trends and counter-trends. From this data we extract and discuss different patterns such as global events, weekly recurring events, regional patterns and outlier events.

Possible use for development:

The visual analytics methods are implemented in a prototype and applied to the provided data to enable and support users in the discovery of global and local patterns, outliers, trends, counter-trends, periodicity and anomalies. The insights gained in the exploration and analysis process can be used for better policy decision making.



Temporal matrix (right) with annotated events (left). A more detailed description can be found in the paper.

Main results:

- We developed a highly interactive prototype system for the exploration of massive mobile phone data in context of the D4D challenge. Using a visual matrix we provide and discuss techniques for the discovery of patterns. We found amongst others:
  - Increase in number of calls correlating with local and global religious events, such as Pilgrimage to *Touba* and *Popenguine* and the end of Ramadan and the Feast of Sacrifice.
  - Towers activated or deactivated throughout the year.
  - Week-weekend patterns for the identification of commercial areas.
  - Identification of Islamic and Christian areas.
  - Correlations of increased call intensity with the harvesting season and weather conditions influencing call intensity such as thunderstorms.

Methods:

- We choose a visual matrix as starting point for the exploration process due to its flexibility and scalability. Furthermore, we provide a multiple coordinated view solution with linked geographic and temporal views. The most important features are:
  - Providing flexibility of attribute projection on both axes.
  - Color-mapping.
  - Hierarchical aggregation
  - Normalization and clustering
  - Summarizing histograms
  - Interaction
  - Coupling with other visualizations.

Full paper is here:

http://tinyurl.com/d4d2014reloaded

Data sources used for this project:

- D4D data set 1, com between antenna [x]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: International Disaster Database — Source: http://www.emdat.be/database
- Type of data: historical weather data — Source: http://wunderground.com/history
- Type of data: various online news media — Source: internet

Main Tools used:

- Qt/C++
- Various algorithms (clustering, normalization)

Open Code available:

- Yes [ ]
- No [x]

Adaptive Power Load Balancing in Cellular Networks

Timothy Larock, Xiaojun Feng, Mariya Zheleva, Petko Bogdanov
Department of Computer Science, University at Albany SUNY

Motivation

- Projections on Internet traffic demand indicate that the traffic in 2018 alone will be as much as the sum traffic from 1984 to 2013.
  - Majority will be handled by mobile networks.
- Mobile networks demand is often unevenly distributed in space.
  - Spatial regularization through load balancing.
  - Current methods are *reactive*, risking deteriorated user experience
- We design *proactive* load balancing techniques based on manipulation of emitted power.
  - Model the network coverage as a power diagram, a generalization of voronoi diagrams.
  - Evaluate the feasibility of this approach on a real-world network: Orange™ in Senegal.

Contributions

- Using real-world data from Senegal, we demonstrate the existence of **significant disparity of load across towers** in a cellular network over time.
- We introduce a novel approach based on power diagrams for **proactive re-distribution of users**, designed to minimize cell tower overload and maximize utilization.
- We perform extensive evaluation of our spatial load balancing approach and demonstrate that it has the **potential of improving the operation of the existing network** in Senegal as a concrete example.
- We provide an **extensive discussion** of the implication of our approach to both commercial cellular networks, but also ones deployed in remote areas and for the purposes of disaster relief.

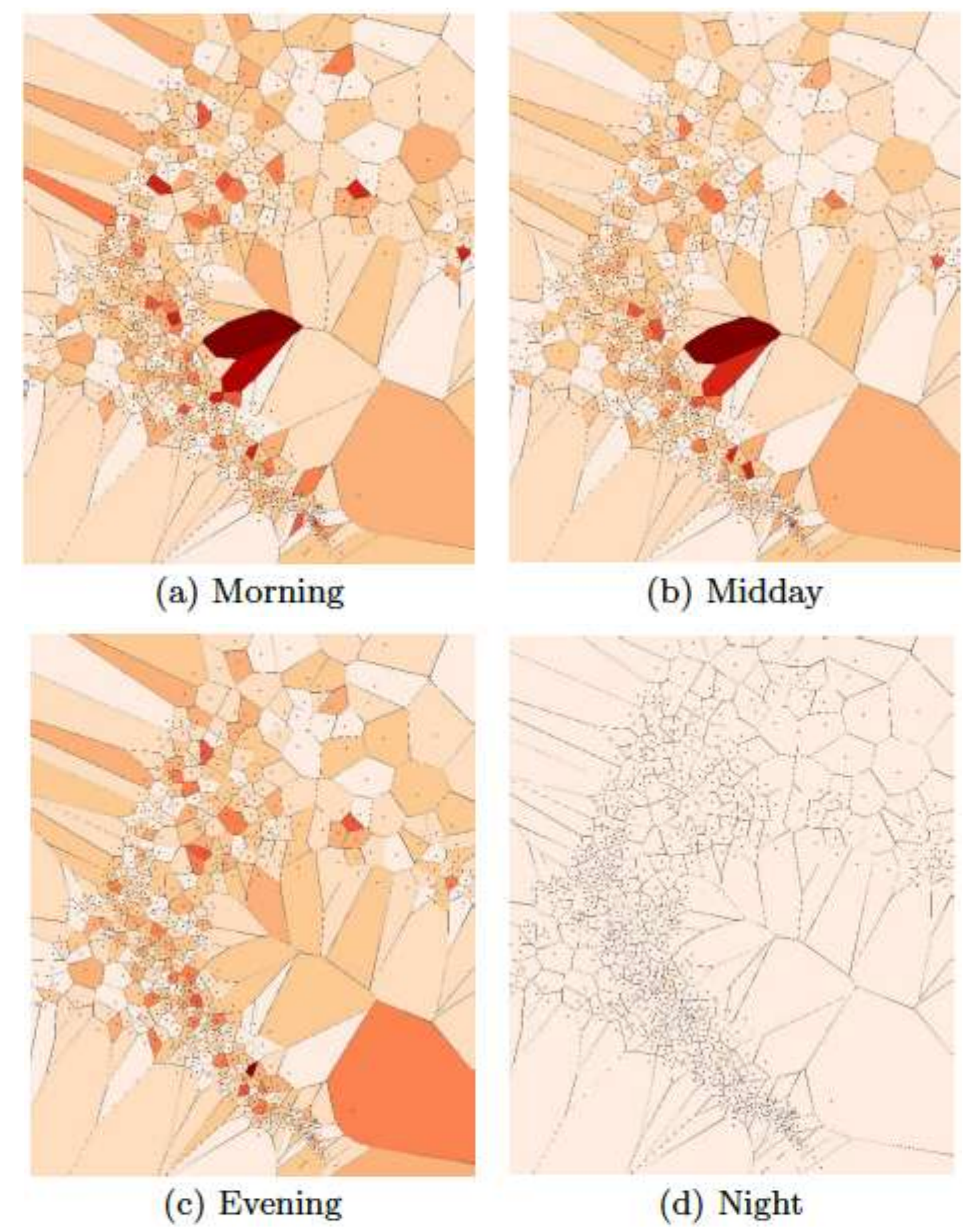
Discussion

- Implications on commercial cellular network deployments.
- Opportunities in community cellular networks in rural or disaster areas.

Data

Our analysis is based on cellular network traces provided by Orange™ in Senegal for the D4D Challenge.

- **Orange™ dataset**
  - High level antenna information collected over the month of January 2013.
  - Perturbed antenna location information.



Voronoi diagrams showing cells in the region of Dakar during four periods of the day. The diagrams are colored based on the average number of calls in each cell during each time period compared to the maximum average call value observed thus far. Darker red color in a cell corresponds to a higher number of calls.

It is evident that there are multiple instances of neighboring cells of different loads.

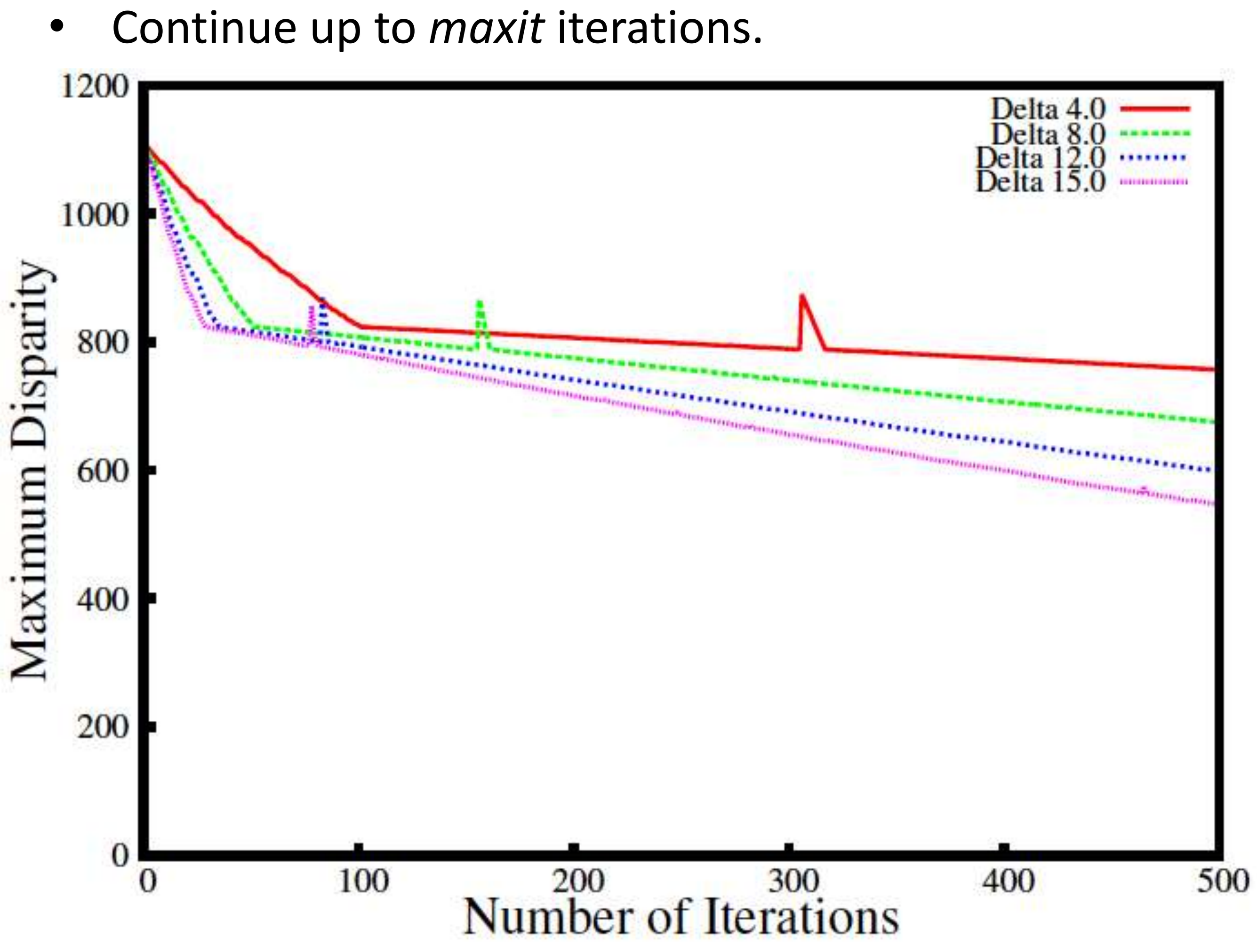
Adaptive Power Assignment

Intuitively, our Adaptive Power Assignment (APA) algorithm identifies cells whose load is very different than that of its neighbors (dubbed high **discrepancy** cells) and updates their power so that the load spreads within the neighborhood.

Discrepancy: $$\delta_i = \frac{\sum_{j \in N_i} |V_i - V_j|}{|N_i|}$$

Algorithm:
Input: Set of N sites, site loads V, atomic unit of power increment Δ
Output: Optimal power distribution

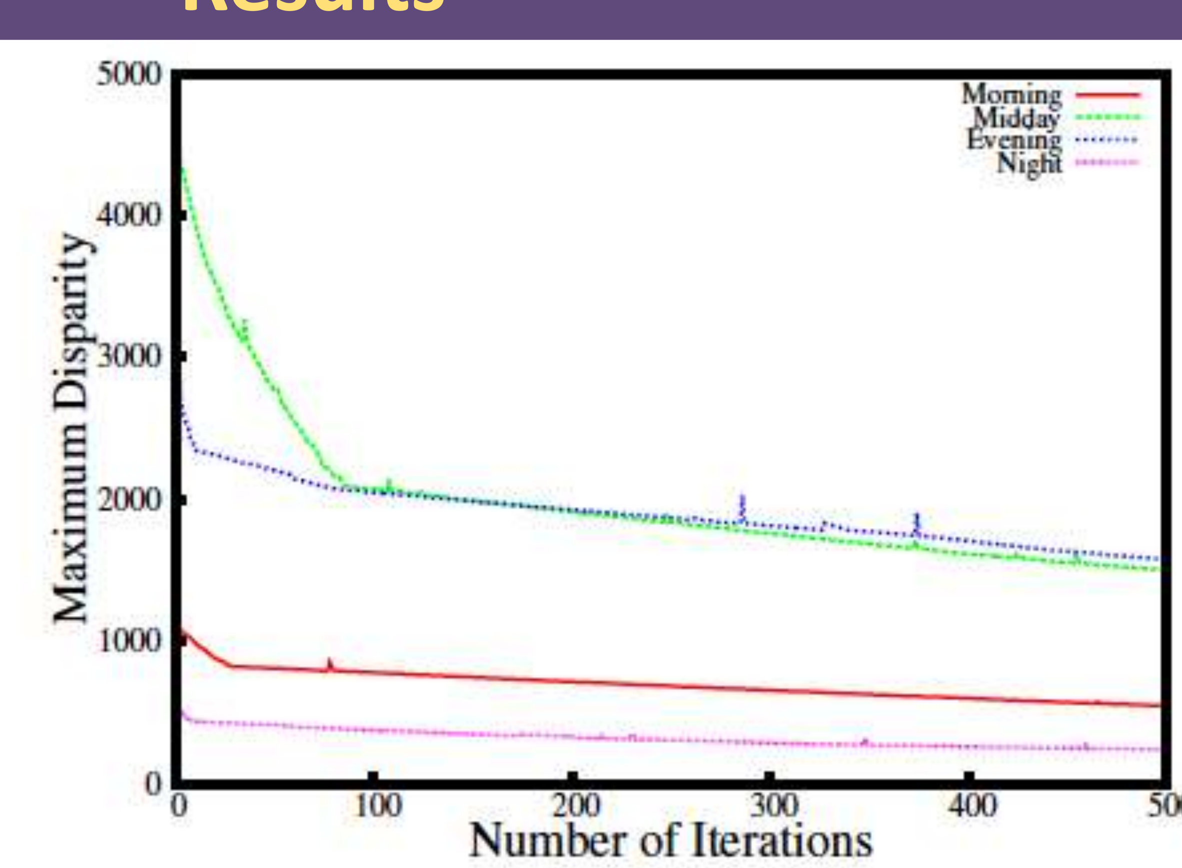
- Find the highest discrepancy site
- Compute an updated power diagram
- Continue up to *maxit* iterations.

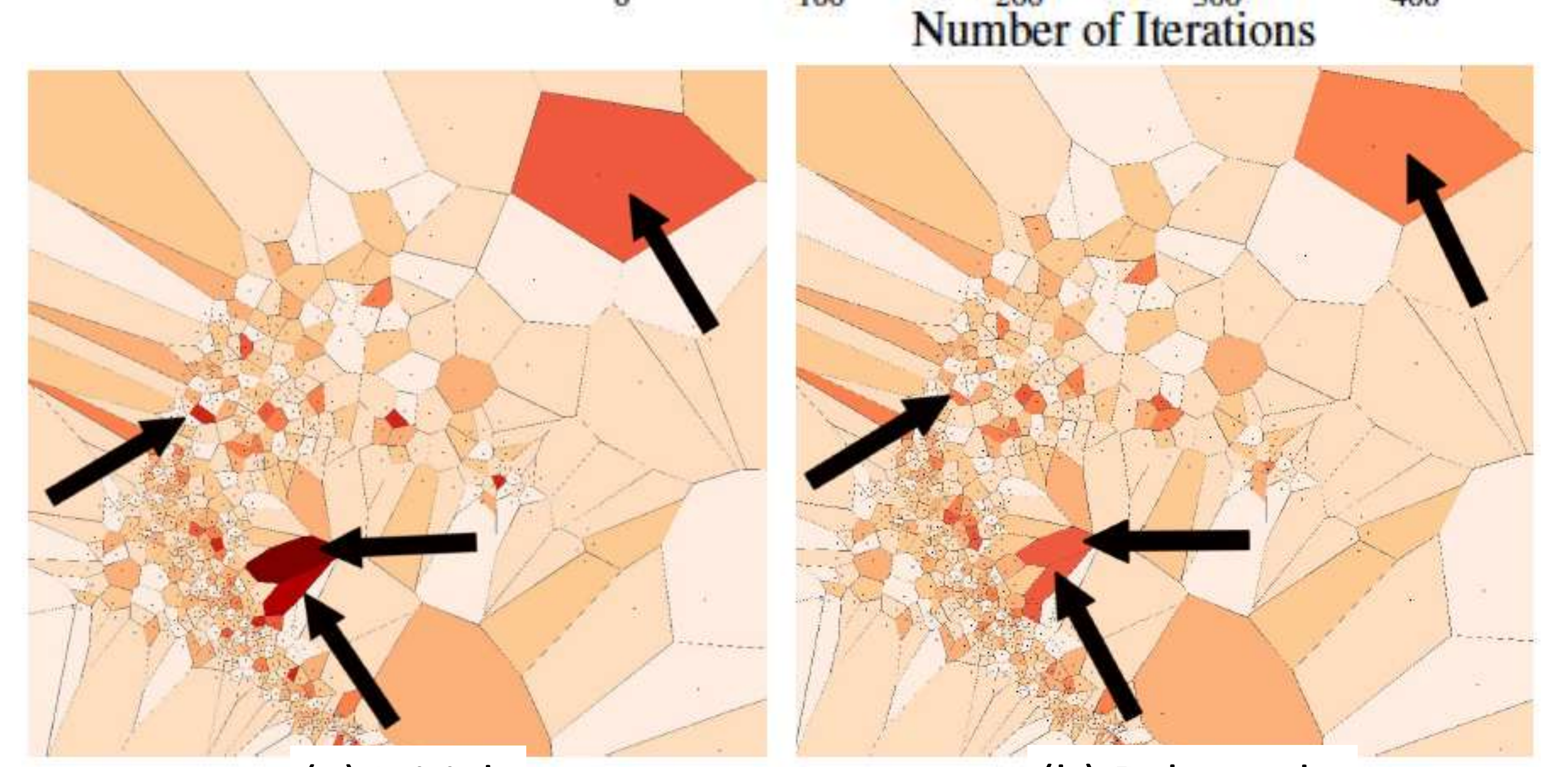


Change in maximum disparity in average calls over 500 iterations of the APA algorithm.

Results

Change in maximum disparity in average calls over 500 iterations at different times of the day, each using delta 15.0.





(a) Initial (b) Balanced

(a) Initial Voronoi diagram representing the average call load of sites in the Dakar region in January 2013. (b) Power diagram representing the same load and sites after 500 iterations of the greedy APA algorithm with Δ=8.0