Earth mover's distance

Jiajie Shi

EMDE and NEMD

We have two variations of EMD ,respectively, NEMD (Normalized Earth Mover's Distance) and EMDE (Earth Mover's Distance with Evaporation). They emphasize different assumptions regarding the change in mobile usage patterns. NEMD assumes that the usage among different tiles changes simultaneously, suggesting a collective shift in user behavior across the entire area. On the other hand, EMDE places emphasis on the idea that changes in usage among tiles are independent, and the amount of evaporation can reflect the change of usage per tile.

In a more generalized approach, it's possible to assign certain weights to the evaporation, which could capture a broader range of scenarios in user behavior changes. In this work we only consider EMDE and NEMD.

Experiments

To calculating EMDE and NEMD for whole tiles in Paris, we initially reduce the data granularity by a factor of 100. In another words, we consolidating the usage of neighboring 100 tiles into a single tile, thereby alleviating computational cost. In terms of flow weights, we determine each flow's weight based on the physical distance between tiles. Specifically, we calculate the euclidean distance from the center of one tile to the center of another tile. In the case of NEMD, an additional step is implemented to normalize the usage of each state. This normalization process ensures that the total usage between the two states where the flow is calculated remains the same.

Two plots below illustrate Instagram usage patterns on March 19th between 9:00 and 9:15 am. In the first EMDE graph, we observe flows originating from both the north and southeast regions converging towards the center of Paris. On another head, the second NEMD graph reveals a somewhat distinct pattern. While some flows still converge at the center of Paris, there is a noticeable decrease in usage in the northeastern area. (the green blocks mean decrease at the tile while red ones means increase).

The comparison between the two versions of EMD is quite evident. NEMD primarily reflects the flow of population or users, showcasing how people move and distribute themselves over a given area. In contrast, EMDE offers information related to the activation or deactivation of mobile apps. It helps us

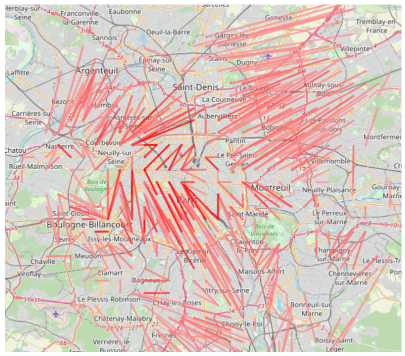

Figure 1: NEMD flows of of Instagram in Paris on March 18th 9:00-9:15 AM

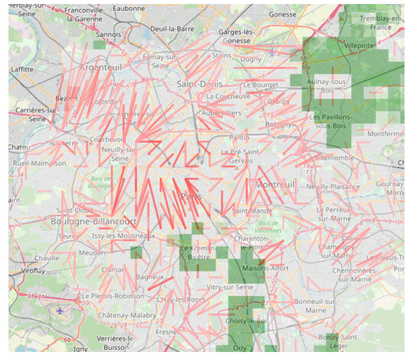

Figure 2: EMDE flows of Instagram in Paris on March 18th 9:00-9:15 AM

understand how apps are used or switched on and off by users within the same spatial context. These two EMD variants provide distinct insights into user behavior and mobile application usage patterns.

Another illustrative example involves running the algorithm locally to observe Instagram usage changes around Prince Park Stadium during a football match scheduled for April 3rd. In this case we don't need to coarse the data. The algorithm is applied to 400 tiles around the stadium. Below are the population flow patterns observed from 8:00-8:15 pm, an hour prior to the start of the match. Both EMDE and NEMD clearly depict the movement of crowds towards the stadium, as indicated by the arrows converging from the periphery towards the center.

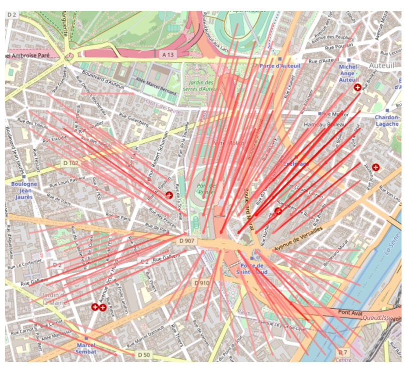

Figure 3: NEMD flows of Instagram near Princes Park on April 3rd 8:00-8:15 pm

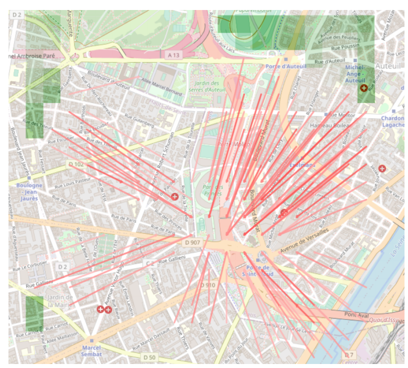

Figure 4: EMDE flows of Instagram near Princes Park on April 3rd 8:00-8:15 pm

In addition to examining flow patterns in isolation, another approach for identifying anomalies is to compare flow data at a specific time with data from a typical or usual time interval. Below we present EMDE flows for April 15th near Notre Dame, where a fire occurred near 6:20 pm. To detect anomalies, we compared the flows with the flow of the same time interval on April 8th, which is a week before the fire incident. This comparative analysis can help find irregularities in the flow patterns. Despite the overall increase in Instagram usage during both time intervals, the flow patterns differ significantly. On April 8th, there are observable flows of users moving from nearby tiles to tiles near Notre Dame, suggesting a typical influx of visitors or activity in that area. Conversely, on April 15th, after the fire incident, the flow dynamics change noticeably. Flows stop going to tiles near Notre Dame, indicating a reduction in visitors in that area. Instead, Instagram usage increases autonomously on tiles near Notre Dame. This shift in flow patterns could be caused by the fact that people who were not in the vicinity of Notre Dame cancelled their plan going there, while those who were already near the site turned to Instagram to share updates or information about the accident. This analysis underscores how changes in flow patterns can offer insights into the behavioral responses of individuals to significant events.

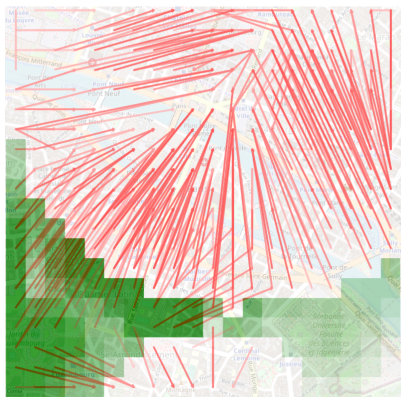

Figure 5: EMDE flows of Instagram near Notre Dame on April 8th 6:15-6:30 pm

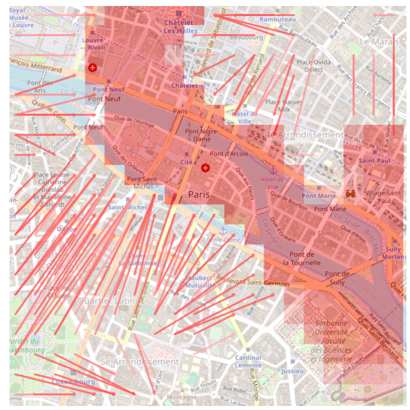

Figure 6: EMDE flows of Instagram near Notre Dame on April 15th 6:15-6:30 pm