VoroGraph: Visualization Tools for Epidemic Analysis

Cody Dunne
IBM Watson
1 Rogers St.
Cambridge, MA 02141
cdunne@us.ibm.com

Mauro Martino
IBM Watson
1 Rogers St.
Cambridge, MA 02141
mmartino@us.ibm.com

Michael Muller
IBM Research
1 Rogers St.
Cambridge, MA 02141
michael_muller@us.ibm.com

Nicola Perra
Northeastern University
Nightingale Hall, 132 C
Boston, MA 02115
n.perra@neu.edu

Abstract

Epidemiologists struggle to integrate complex
information about the incidence and spread of disease,
in relation to population density and other demographic
conditions, at geographical scales ranging from global
air travel down to local commuting. A partial solution
overlays air travel as arcs above color-coded maps.
However, commuting is not shown and it is often
challenging to understand changing relationships due
to the visual complexity arcs introduce. Moreover,
when region sizes and shapes vary their color-codings
become difficult to perceive. We introduce three
visualizations which combine representations of
population, movement, and disease spread at a local
scale that is consistent with a zoomable global scale:
(1) a map with commuting border encodings, (2) a
centroidal Voronoi tessellation morphing technique,
and (3) a meta-layout showing commuting alongside
air travel. Our work provides mid-level abstractions that
expert epidemiologists can use for insights into
contagion.

Author Keywords

Epidemiology; Emergency Management; Visualization;
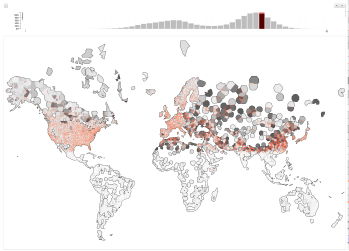Graph/Network Data; Geographic/Geospatial Data
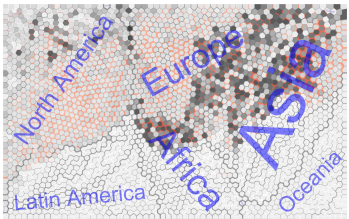
Figure 1: Border-encoded map and timeline.

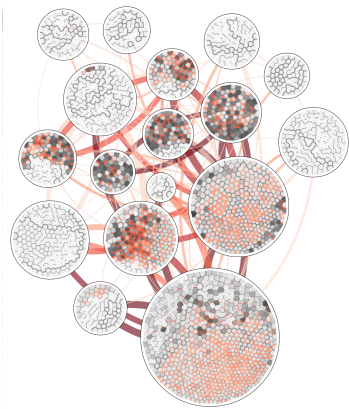**Figure 2:** Centroidal Voronoi tessellation (CVT) visualization.

Figure 3: CVT-based meta-layout.

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction

Consider a new strain of influenza starting from New York City in mid February.[1] New York is densely populated, and further the large majority of the population would be susceptible to the virus. The disease could spread quickly reaching pandemic proportion. Indeed, the millions of commuters, national, and international visitors could carry the virus home with them. As in Mexico in 2009, the authorities would have to face both the local and global spreading of the disease.

The global epidemic and mobility (GLEAM) model [1] is designed to simulate and predict these events providing support and intelligence to decision makers. However, available visualization tools such as GLEAMviz [1] typically used to share predictions and risk maps with authorities or used by modelers to explore the simulations results, focus just on long-range spreading patterns. The inherent complexity of these relationships motivated its designers to only show the first infected flight on any route, rather than showing the route in all time slices. They do not show local diffusions that nonetheless drive the unfolding of the disease across nearby cities. This strongly limits our understanding of the complete phenomena. Traditionally for epidemic visualization, only long-range travel is shown.

Networks such as these diffusion patterns are often difficult to comprehend, even with advanced

---

[1] The 2009 H1N1 pandemic started in mid Feb. near Mexico City.

visualization techniques. For geographic data, we are further limited in that spatial layout cannot be easily used to show relationships between regions. Local relationships between contiguous geographies are especially hard to visualize, as many of these geographies are only tiny features on a standard world map. Ideally we would use visual encodings like color to show quantitative properties such as the number of infected in a region. However, color can be challenging to perceive when the shape and area of the regions varies substantially. Another layer of complexity is that the relationships between regions are evolving.

We have developed a web-based visualization tool called VoroGraph which directly tackles such limits, allowing epidemiologists to investigate and present, for the first time, both local and long range transmission patterns. As visualization researchers, before we designed the interface we first worked with epidemiologists to understand the challenges they face. We address these challenges with three novel visualization techniques which we show in Figs. 1 to 3:

- A map which uses visual encodings on borders between neighboring regions to show local relationships;

- A centroidal Voronoi tessellation (CVT) [2] visualization which morphs the map in a space-filling manner while preserving relative positions, so as to highlight region properties as well as local relationships; and

- A CVT-based meta-layout which displays aggregate long-distance in addition to local relationships.
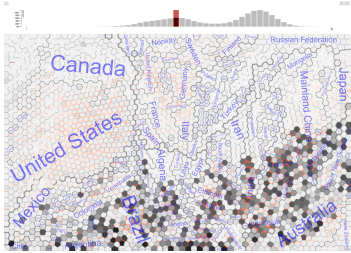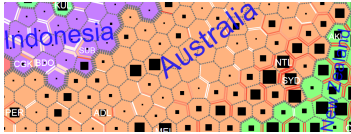
Figure 4: CVT morphing of the map.

Figure 5: Closeup of the CVT view showing directed commuting travel between basins. For example, see the asymmetric commuting borders of SUB in Indonesia. Data squares are shown for basin infection ratio, with categorical coloring by country and labels for countries and top airports.
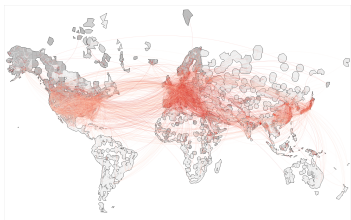

Figure 6: Border-encoded map with flight non-contiguous relationships drawn as arcs at 20% opacity.

Design

Here we use the term **basin** to describe a geographic area from the GLEAM model containing an airport.

Border-Encoded Map

Commuting relationships in GLEAM exist only between contiguous basins on a map. Thus, a meaningful extension to any map is to encode these relationships on shared borders using size and color. We have done this for a Mercator projection, shown in Fig. 1, where regions are colored light to dark gray depending on the ratio of infected population. Border size encodes the total number of commuters between contiguous basins, while the color on a white to red scale shows how many of them are infected.

Centroidal Voronoi Tessellation (CVT) Morphing

To address the border-encoded map limitations we investigated morphing the map using a centroidal Voronoi tessellation (CVT) (Figs. 2 and 4). CVTs can be iteratively computed using Lloyd's algorithm [4]. Optimal CVTs, which can be approximated in practices, have several properties that make them effective information visualization tools:

- Cells are evenly distributed and space-filling,
- Cells can are equal area,
- Relative positions are preserved,
- Cells are generally regular hexagons,
- Cells have an aspect ratio close to one, and
- Cell positions and shapes are deterministic.

Labeling

With the additional basin area in the CVT view, we can now add labels for continents, regions, countries, or airports (Figs. 2, 4 and 5). We first find the proper aspect ratio to display the chosen font and label. We then union all the basins of a target area and find the largest rectangle at any angle which fits within it. This assures we have the largest label desired. We aslo show the most populated 5% of basins in each country.

Border Encodings

We draw two offset parallel borders wherever people commute, and the border closest to a basin shows outbound travel while the further border shows inbound (Fig. 5). The borders are sized by the total number of commuters traveling in that direction and colored on a white-red scale by the ratio infected. If neighboring basins belong to different countries, regions, or continents an appropriately sized fence is drawn to designate the hierarchy. In Fig. 5 we can see the separation between Asia (top left) and Oceanea (bottom right) shown with a wide fence. Individual countries are delineated using thin fences.

Rather than fences, or in addition to them as in Fig. 5, we can use categorical coloring for the hierarchy. However, this makes understanding a color coding for basin properties more challenging. Thus we provide the option of using data squares sized according to the infected population. We can see that Juanda International Airport (SUB) in Indonesia has few infected (small square), though has a highly infected neighbor (large square).

CVT Meta-Layout

Fig. 6 shows the effect of drawing the flights as arcs on top of the border-encoded map and CVT views. Even

with the arcs at 20% opacity, it is difficult to make sense of the map view. Instead, we create an aggregate meta-network using the geographic hierarchy, where meta-nodes represent groupings of basins and the meta-edges between them represent the aggregate flights cutting across the hierarchy. This meta-network can be laid out using any force-directed layout that is aware of edge weights, for example to scale meta-edge distance inversely with the total infected flying population between two meta-nodes. Fig. 3 shows meta-edges are drawn as arcs colored by the ratio of infected population flying and arc width signifies the total population flying. The nested CVT basins are still larger and better for encoding properties and commuting relationships than the map view.

Animated Transitions
When changing between views it is important not to disrupt the mental map of the user, especially if they are unfamiliar with the new layout. Animated and staggered transitions can help users understand the change between views [3]. On a view change, each basin is morphed using path tweening to its new shape and translated to its new location, and the animations are staggered according to the continent hierarchy.

Timeline
As the evolution of the global network properties are important to understand, we included a timeline widget visible along the top of Figs. 1 and 4. It shows the total number of infected people over time, and allows the user to select any timestep or play through time to update the other views.

Conclusion
Epidemiologists often need to look at both local commuting and long-distance flight relationships to understand the spread of a disease. We developed several visualizations to assist in their analyses and presentations, including a border-encoded map and two views based on centroidal Voronoi tessellations.

We conducted a first formative evaluation of the three visualizations with nine epidemiologists, and any of the issues of disorientation appear to be addressable through simple re-designs. Some made concrete design recommendations regarding labels, colors, boundary-treatments, and interaction mechanisms. We are continuing to work with them as they want to incorporate an improved version of VoroGraph as part of the publicly available GLEAMviz suite [1].

There are many other domains that would benefit from these techniques, particularly those where it is important to explore the interplay between local and global relationships.

References

1. Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., and Vespignani, A. Multiscale mobility networks and the spatial spreading of infectious diseases. *PNAS*, 106(51), 2009: 21484–21489.
2. Du, Q., Faber, V., and Gunzburger, M. Centroidal Voronoi tessellations: applications and algorithms. *SIAM Review*, 41(4), 1999: 637–676.
3. Heer, J., and Robertson, G. G. Animated transitions in statistical data graphics. *TVCG*, 13(6), 2007: 1240–1247.
4. Lloyd, S. P. Least squares quantization in PCM. *ToIT*, 28(2), 1982: 129–137.