Spatial Partitioning Algorithms for Data Visualization

Raghuveer Devulapalli[a], Mikael Quist[a] and John Gunnar Carlsson[a]

[a]Industrial and Systems Engineering, University of Minnesota, Minneapolis, USA

ABSTRACT

Spatial partitions of an information space are frequently used for data visualization. Weighted Voronoi diagrams are among the most popular ways of dividing a space into partitions. However, the problem of computing such a partition efficiently can be challenging. For example, a natural objective is to select the weights so as to force each Voronoi region to take on a pre-defined area, which might represent the relevance or market share of an informational object. In this paper, we present an easy and fast algorithm to compute these weights of the Voronoi diagrams. Unlike previous approaches whose convergence properties are not well-understood, we give a formulation to the problem based on convex optimization with excellent performance guarantees in theory and practice. We also show how our technique can be used to control the shape of these partitions. More specifically we show how to convert undesirable skinny and long regions into fat regions while maintaining the areas of the partitions. As an application, we use these to visualize the amount of website traffic for the top 101 websites.

Keywords: Data visualization, Space partitions, Weighted Voronoi diagrams

1. INTRODUCTION

Dividing a given geographic region into sub-regions in an optimal way is a natural problem that belongs to many different domains, such as air traffic control,[1] congressional districting,[2] vehicle routing,[3] facility location,[4] and urban planning.[5,6] A canonical method for partitioning a region is the *Voronoi diagram*, in which the region is partitioned into smaller sub-regions based on proximity to a set of "landmark" points, as shown in Figure 1. Simply put, given a geographic region R containing a set of landmark points $\{p_1, \ldots, p_n\}$ the Voronoi cell associated with p_i, denoted V_i, is defined as the set of points $x \in R$ where $\|x - p_i\| \leq \|x - p_j\|$ for all indices j:

$$V_i = \{x \in R : \|x - p_i\| \leq \|x - p_j\| \, \forall j\} \ ,$$

where $\|\cdot\|$ most commonly denotes the Euclidean norm. Such partitions have been generalized in a number of ways by introducing a weight vector $\mathbf{w} = (w_1, \ldots, w_n)$ associated with the landmark points that controls the

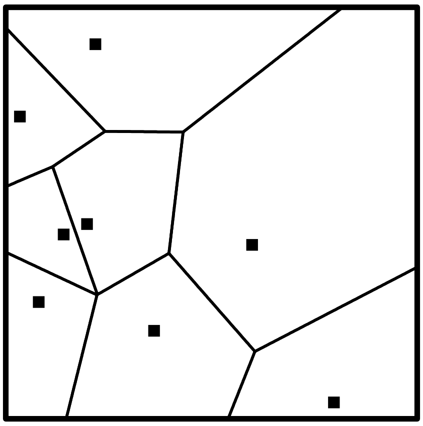


Figure 1: A Voronoi diagram of $n = 8$ points in the unit square.

Further author information: (Send correspondence to Raghuveer Devulapalli)
Raghuveer Devulapalli: Homepage: http://www.tc.umn.edu/ devul002/, E-mail: devul002@umn.edu
John Gunnar Carlsson: Homepage: http://menet.umn.edu/ jgc/, E-mail: jgc@umn.edu

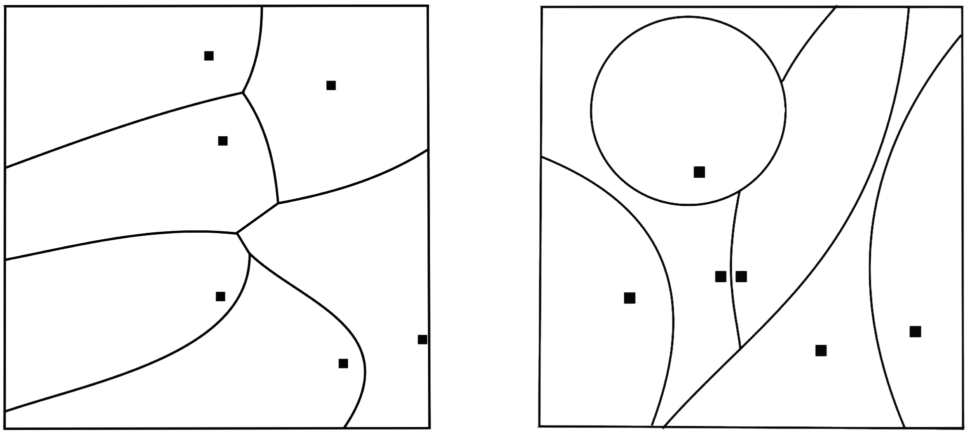


Figure 2: Additive and multiplicatively weighted Voronoi diagrams in the unit square.

shapes and sizes of the regions. Two common such generalizations are the *additively-weighted Voronoi diagram*, in which we define Voronoi cells as

$$V_i = \{x \in R : \|x - p_i\| - w_i \leq \|x - p_j\| - w_j \, \forall j\} \ ,$$

and the *multiplicatively-weighted Voronoi diagram*, in which we have

$$V_i = \left\{x \in R : \frac{\|x - p_i\|}{w_i} \leq \frac{\|x - p_j\|}{w_j} \, \forall j\right\} \ ,$$

which are shown in Figure 2. Observe that in both of these generalizations, the size of cell V_i increases as w_i increases. It is also not hard to show that the boundaries between adjacent Voronoi cells V_i and V_j are hyperbolic arcs in the additive model and circular arcs in the multiplicative model.

In this paper, we consider the problem of partitioning an abstract information space, as opposed to a physical region as mentioned in the beginning of this section; similar attempts to ours include.[7–13] Specifically, our objective is to find an effective representation of an information space, such as a set of documents, as a planar diagram that conveys relevant information. In this setting, each document is represented as a region in the plane, in the same way as is shown in Figure 3. There are two major objectives that should be considered in designing such a diagram effectively: first, documents containing similar content should be placed in close geographic proximity to one another. Second, documents with larger significance or relevance should be represented by regions that are larger than those corresponding to documents with less significance.

Being based on the landmark points $\{p_1, \ldots, p_n\}$, the Voronoi framework lends itself well to the first objective described; that is, given a set of n documents and additional information regarding their relationships to one another, one can place the landmark points in a way that is commensurate with the relationships between the documents using graph visualization software such as GraphViz or Gephi (see Figure 4). A less-studied problem is how to leverage the Voronoi framework in service of the second objective. To this end, a problem elegantly posed by Reitsma et al.[15] is as follows: suppose that the landmark points $\{p_1, \ldots, p_n\}$ are given in a region with area 1, together with a set of desired areas $\{A_1, \ldots, A_n\}$ that also sum to 1 (and which implicitly are related to the significance or relevance of the n documents); can one find a weight vector $\mathbf{w}^*$ in the additive or multiplicative model such that $\text{Area}(V_i) = A_i$ for all i? The authors give an affirmative answer and describe an iterative, raster-based scheme for determining such a weight vector under the multiplicative model. The main drawbacks to this scheme relate to algorithmic efficiency, both in a practical and theoretical sense: as the algorithm is based on a form of fixed-point iteration, there is little in the way of performance guarantees, and consequently, the proposed methodology does not scale well as the problem size becomes large.

In this paper, we give a fast algorithm for finding the desired weight vector $\mathbf{w}^*$ in either the additive or multiplicative model. Rather than using fixed-point iteration, our approach is based on principles from variational calculus, specifically duality theory in linear programming over infinite-dimensional vector spaces, and thus

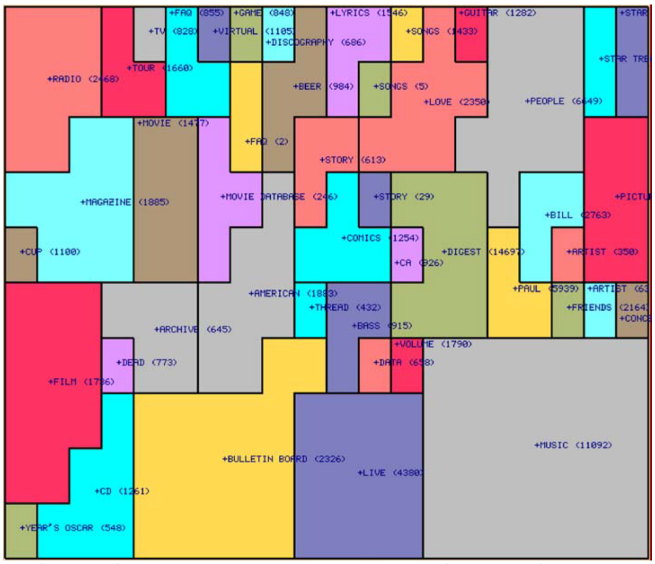
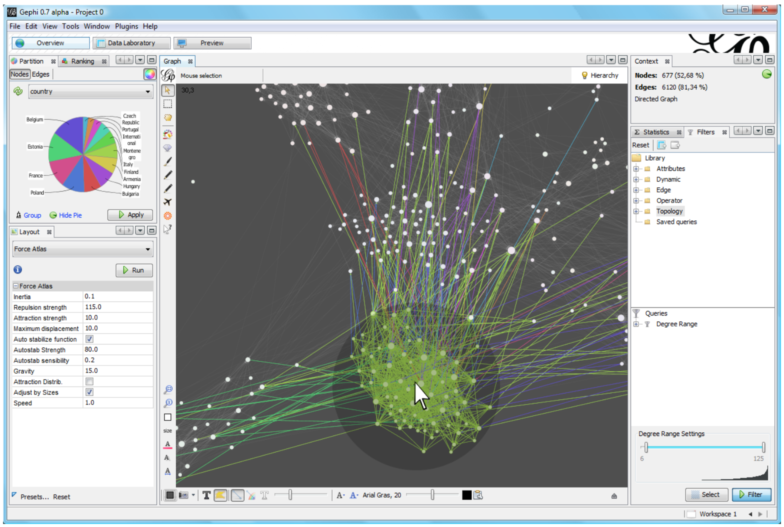


Figure 3: A planar map of an information space consisting of a set of documents, as constructed in.[14]



Nodes: 677 (52,68 %)
Edges: 6120 (81,34 %)
Directed Graph

Force Atlas	
Inertia	0.1
Repulsion strength	115.0
Attraction strength	10.0
Maximum displacement	10.0
Auto stabilize function	☑
Autostab Strength	80.0
Autostab sensibility	0.2
Gravity	15.0
Attraction Distrib.	☐
Adjust by Sizes	☑
Speed	1.0

Figure 4: A graph visualization generated by Gephi.

inherits excellent theoretical and practical performance guarantees. We further show how to apply a "homotopy method" to enable better control over Voronoi regions V_i. Finally, we demonstrate the effectiveness of our algorithm in various computational experiments applied to a list of major internet sites.

2. PROCEDURE

In this section, we consider the problem of finding the weight vector $\mathbf{w}^*$ as in the preceding section, assuming that the landmark points $\{p_1, \ldots, p_n\}$ are already fixed in a convex planar domain R of area 1 (which will typically be a rectangle or square in an informational display) and that the desired areas of sub-regions $\{A_1, \ldots, A_n\}$ are given. To introduce our algorithm, we find it useful to temporarily disregard the Voronoi framework and simply consider useful properties that the sub-regions $R_1, \ldots, R_n$ associated with the landmark points should have. Suppose that region R_i is associated with point p_i. Because we presumably located point p_i in a useful way relative to the other landmark points, it is natural to desire that the region R_i should be "close" to p_i as well, so as to inherit its strategic placement. This "closeness" can be measured by what we call the *Mean Distance Function* (MDF) defined by

$$\mathrm{MDF}(p_i, R_i) := \iint_{R_i} \|x - p_i\| \, dx \,,$$

which is simply proportional to the mean distance between a point $x \in R_i$ selected uniformly at random and the landmark point p_i. We can therefore consider the problem of designing an optimal partition $\{R_1, \ldots, R_n\}$ that minimizes the above function, subject to the constraints on the areas, expressed as the following optimization problem (which, we re-iterate, does not involve the Voronoi framework just yet):

$$\begin{aligned} \mathrm{minimize}_{R_1,\ldots,R_n} \sum_{i=1}^{n} \mathrm{MDF}(p_i, R_i) & \qquad s.t. \\ \mathrm{Area}(R_i) &= A_i \; \forall i \\ R_i \cap R_j &= \emptyset \; \forall i,j \\ \bigcup_{i=1}^{n} R_i &= R\,. \end{aligned} \tag{1}$$

It turns out that the optimal solution to (1) can be recovered by solving an n-dimensional *convex optimization problem*, that is, an optimization problem whose objective function and feasible set are both convex, as made concrete by the following theorem:

THEOREM 2.1. *The sub-regions $\{R_1^*, \ldots, R_n^*\}$ that solve problem (1) can be recovered by solving the following optimization problem:*

$$\begin{aligned} \mathrm{maximize}_{\boldsymbol{\lambda}} \iint_R \min\{\|x - p_i\| - \lambda_i\} \, dx & \qquad s.t. \\ \sum_{i=1}^{n} \lambda_i &= 0\,. \end{aligned} \tag{2}$$

Specifically, if $\boldsymbol{\lambda}^*$ denotes the optimal solution to problem (2), then sub-region R_i^* consists of those points $x \in R$ for which $\|x - p_i\| - \lambda_i^*$ is minimal:

$$R_i^* = \left\{ x \in R : \|x - p_i\| - \lambda_i^* \le \|x - p_j\| - \lambda_j^* \, \forall j \right\} \,.$$

Proof. Problem (1) can be written as an *infinite-dimensional integer program* in which the optimization variables are indicator functions $\mathcal{I}_i(\cdot)$ rather than regions R_i. Specifically, if we let $\mathcal{I}_i(x)$ denote a $\{0, 1\}$ function that indicates whether or not point $x \in R$ belongs to sub-region R_i, we can re-write (1) as

$$\begin{aligned}
\text{minimize}_{\mathcal{I}_1(\cdot),\ldots,\mathcal{I}_n(\cdot)} \iint_R \|x - p_i\| \mathcal{I}_i(x)\, dx \qquad & s.t. \\
\iint_R \mathcal{I}_i(x)\, dx \quad &= \quad A_i\ \forall i \\
\sum_{i=1}^{n} \mathcal{I}_i(x) \quad &= \quad 1\ \forall x \in R \\
\mathcal{I}_i(x) \quad &\in \quad \{0,1\}\ \forall i, x\,.
\end{aligned} \tag{3}$$

If we then relax the integrality constraint, we obtain an infinite-dimensional linear program:

$$\begin{aligned}
\text{minimize}_{\mathcal{I}_1(\cdot),\ldots,\mathcal{I}_n(\cdot)} \iint_R \|x - p_i\| \mathcal{I}_i(x)\, dx \qquad & s.t. \\
\iint_R \mathcal{I}_i(x)\, dx \quad &= \quad A_i\ \forall i \\
\sum_{i=1}^{n} \mathcal{I}_i(x) \quad &= \quad 1\ \forall x \in R \\
\mathcal{I}_i(x) \quad &\geq \quad 0\ \forall i, x\,.
\end{aligned} \tag{4}$$

By applying standard results of vector space optimization (see Theorem 1 of[16]), we can show that problem (2) is the *dual* of Problem (4). The characterization of the optimal regions R_i^* in terms of $\boldsymbol{\lambda}^*$ is precisely the complementary slackness conditions of problems (2) and (4). □

The following corollary is immediate:

THEOREM 2.2. *The optimal solution to (1) is an additively weighted Voronoi diagram in which* $\text{Area}(V_i) = A_i$ *for all i.* We can thus see that an additively weighted Voronoi diagram with appropriate areas $\{A_1, \ldots, A_n\}$ can be obtained easily once we have solved problem (2). Thus, it will suffice to turn our attention to the issue of how to solve (2) efficiently.

It is not hard to verify using standard methods that the objective function of problem (2) is concave and differentiable as a function of $\boldsymbol{\lambda}$. In particular, it can be shown (see Section 4.1 of[16]) that

$$\frac{\partial}{\partial \lambda_i} \left(\iint_R \min\{\|x - p_i\| - \lambda_i\}\, dx \right) = -\text{Area}(R_i)\,,$$

where R_i denotes the set of points $x \in R$ where $\|x - p_i\| - \lambda_i$ is minimal:

$$R_i = \{x \in R : \|x - p_i\| - \lambda_i \leq \|x - p_j\| - \lambda_j\ \forall j\}\ .$$

We can thereby see that (2) is a convex optimization problem (since we are *maximizing* a concave function on a convex set) for which gradient vectors are easy to compute. We can therefore determine the optimal vector $\boldsymbol{\lambda}^*$ (and therefore the optimal sub-regions $\{R_1, \ldots, R_n\}$) using, for example, an analytic center cutting plane method, as described in Algorithm 1.

3. VARIATIONS

In Section 2, we introduced Algorithm 1 which obtained an additively-weighted Voronoi diagram with pre-specified areas. We can perform a very minor modification to this procedure to construct a multiplicatively-weighted Voronoi diagram with pre-specified areas. This is accomplished by considering the following optimiza-

```
Input: A convex, planar region $R$, a collection of points $\{p_1,\ldots,p_n\}$, a collection of desired areas $\{A_1,\ldots,A_n\}$
       such that $\sum_i A_i = 1$, and a threshold $\epsilon$.
Output: A partition of $R$ into $n$ regions $R_1,\ldots,R_n$ that solves problem (1) within tolerance $\epsilon$.
Note: this is simply a standard analytic center cutting plane method applied to problem (2).
Define the initial polyhedron by $\Lambda = \left\{\boldsymbol{\lambda} \in \mathbf{R}^n : \sum_{i=1}^n \lambda_i = 1 \text{ and } \|\boldsymbol{\lambda}\|_\infty \leq M\right\}$ for a large threshold $M$;
/* A suitable value of $M$ is to set $M = nd$, where $d$ is the diameter of $R$.     */
while $\mathrm{vol}(\Lambda) > \epsilon$ do
    Let $\boldsymbol{\lambda}^0$ be the analytic center of $\Lambda$;
    for $i \in \{1,\ldots,n\}$ do
        Let $R_i$ denote the sub-region in $R$ for which $\|x - p_i\| - \lambda_i$ is minimal;
    end
    for $i \in \{1,\ldots,n\}$ do
        Set $g_i := -\mathrm{Area}(R_i)$;
    end
    Set $\Lambda := \Lambda \cap \{\boldsymbol{\lambda} : \mathbf{g}^T\boldsymbol{\lambda} \geq \mathbf{g}^T\boldsymbol{\lambda}^0\}$;
end
return $\{R_1,\ldots,R_n\}$;
```

Algorithm 1: Algorithm BestPartition partitions a given region into sub-regions with pre-specified areas.

tion problem, by comparison with (1):

$$\begin{aligned}
\text{minimize}_{R_1,\ldots,R_n} \sum_{i=1}^n \iint_{R_i} \log(\|x - p_i\|)\, dx \qquad & s.t. \qquad (5)\\
\mathrm{Area}(R_i) \quad = \quad & A_i \;\forall i\\
R_i \cap R_j \quad = \quad & \emptyset \;\forall i,j\\
\bigcup_{i=1}^n R_i \quad = \quad & R\,.
\end{aligned}$$

Using the same procedure as in the proof of Theorem 2.1, we can show that the optimal solution to (7) is determined by the following n-dimensional optimization problem:

$$\begin{aligned}
\text{maximize}_{\boldsymbol{\lambda}} \iint_R \min\{\log(\|x - p_i\|) - \lambda_i\}\, dx \qquad & s.t. \qquad (6)\\
\sum_{i=1}^n \lambda_i \quad = \quad & 0\,.
\end{aligned}$$

It turns out that the optimal sub-region R_i^* consists of those points $x \in R$ such that $\log(\|x-p_i\|)-\lambda_i$ is minimal among all indices:

$$R_i^* = \{x \in R : \log(\|x - p_i\|) - \lambda_i \leq \log(\|x - p_j\|) - \lambda_j \;\forall j\}\ .$$

Of course, by exponentiating both sides, we therefore see that

$$R_i^* = \left\{x \in R : \frac{\|x - p_i\|}{e^{-\lambda_i}} \leq \frac{\|x - p_j\|}{e^{-\lambda_j}} \;\forall j\right\}\,,$$

which is precisely a multiplicatively-weighted Voronoi diagram. It can again be shown that the objective function of (6) is concave and differentiable, and that

$$\frac{\partial}{\partial \lambda_i}\left(\iint_R \min\{\log(\|x - p_i\|) - \lambda_i\}\, dx\right) = -\mathrm{Area}(R_i)\,,$$

so that the problem is just as tractable as in the additive case.

3.1 Enforcing fatness of sub-regions

Paritioning a region using additively weighted Voronoi diagrams can have sub-regions which are long and skinny. For example, Figure 5a shows an additively weighted Voronoi diagram with 10 equal area sub-regions. They have a nice property that every landmark point is always guaranteed to be within its assigned sub-region and that all sub-regions are connected, but as mentioned above, the sub-regions can become long and skinny. For regions with a very small fraction of the total area, this becomes even worse. This is clearly undesirable when visualizing data using partitions.

We observed that for multiplicative Voronoi diagrams, the boundaries are circular arcs and hence regions with small areas are always tend to be circular and fat. Therefore, we propose a "homotopy" method in which we combine the objective functions of additive and multiplicative Voronoi diagrams and minimize a weighted combination of the two objective functions.

$$\text{minimize}_{R_1,\ldots,R_n} \sum_{i=1}^{n} \iint_{R_i} (1-\mu)\|x-p_i\| \ + \mu \log(\|x-p_i\|)\, dx \qquad s.t. \tag{7}$$

$$\begin{aligned} \text{Area}(R_i) &= A_i \ \forall i \\ R_i \cap R_j &= \emptyset \ \forall i,j \\ \bigcup_{i=1}^{n} R_i &= R\,. \end{aligned}$$

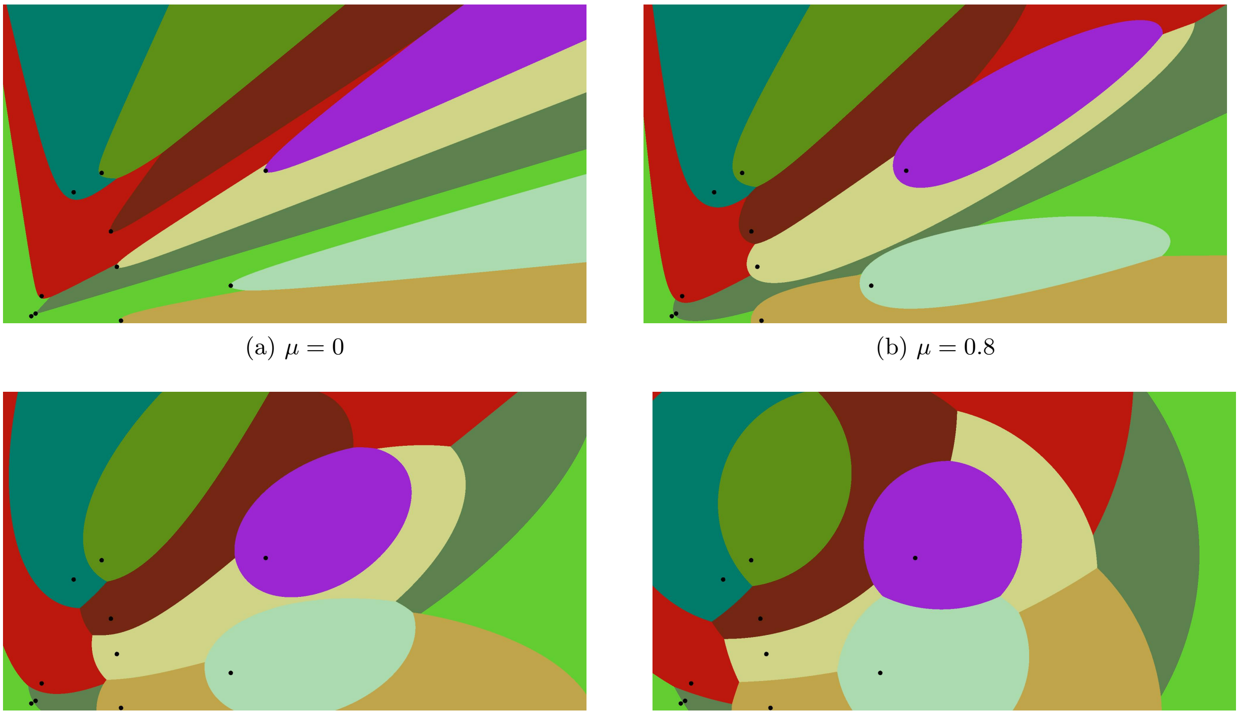
The above problem is similar to an instance of (1) with a modified objective function; we find that, consequently, the optimal sub-regions of this formulation satisfy:

$$R_i^* = \{x \in R : (1-\mu)\|x-p_i\| + \mu \log(\|x-p_i\|) - \lambda_i \leq (1-\mu)\|x-p_j\| + \mu \log(\|x-p_j\|) - \lambda_j \ \forall j\}\ .$$

These sub-regions can be obtained using similar procedures to those discussed previously. Figure 5 shows the effect of this modified objective function. For $\mu = 1$ the partitions correspond to a multiplicative Voronoi diagram, in which sub-regions need not always be connected. But since, for $\mu = 0$ the regions are always connected (because they correspond to an additive Voronoi diagram), it is not hard to show that we can find a "threshold" $\mu^* \in [0,1]$ for which the sub-regions are always connected.

4. INTERNET TRAFFIC VISUALIZATION

In this section, we show the use of Voronoi space partitions to visualize real data. One of the most popular choices for information visualization if the amount of traffic on various internet websites. The information space is represented by a planar rectangle and the landmark points within this rectangle represent the locations of internet websites. Once we partition this space, the size of a sub-region is proportional to the total number of page views on the corresponding website. Using Alexa Internet, Inc. (accessed June 7^{th} 2013), we obtained the amount of website traffic for the top 101 websites between the period September 2007 - June 2013. We first chose a subset of this data consisting of 13 popular websites (mainly social media, search engines and e-commerce) and we embed them on a 2-dimensional planar rectangle. A collection of such "maps" that display the relative "sizes" of these 13 major websites, taken between 2007 and 2013 are shown in Figures 6c to 6e. In these figures, websites with similar content are represented as regions that are in close proximity with one another, such as Facebook and Myspace. The location of the websites are held constants throughout the time period. The partitions are purely additively weighted Voronoi diagrams which produces hyperbolic boundaries. As we can see from the figures, Yahoo! and Myspace were dominant websites in September 2007, but as time progresses we can clearly visualize the decline of Yahoo! and Myspace which are engulfed by Google and Facebook respectively. As of June 2013, Google, Facebook and Youtube together account for more than 85% of the website traffic, relative to these 13 popular websites.



(a) $\mu = 0$

(b) $\mu = 0.8$

(c) $\mu = 0.9$

(d) $\mu = 1$

Figure 5: Equal area partitions for 10 randomly chosen points in the plane. Figure 5a is an additive Voronoi diagram which has long and skinny region. By increasing the weight on the penalty term, we can convert these regions into fat regions. At $\mu = 1$, this is simply a multiplicative Voronoi diagram (Figure 5d)

We now represent the complete data set of 101 websites obtained from Alexa. Based on their genre, these websites were placed close to each other and their locations were obtained by using an open source tool, *Graphviz* (Graph Visualization Software). This is shown in Figure 7a. Based on the traffic information for these websites for June 2013, the map was partitioned using an additively weighted Voronoi diagram. As we can see from Figure 7b, a lot of these regions are long and skinny which is clearly undesirable. We hence, create partitions based on a combination of additive and multiplicative Voronoi diagram using the homotopy method presented above 3.1. A collection of such maps between the time period September 2007 and June 2013 are shown in Figure 8. Because of limited space, only a few major websites have been labeled. A more detailed visualization of the evolution of partitions for top 101 websites are shown in the multimedia file uploaded along with this manuscript.

ACKNOWLEDGMENTS

REFERENCES

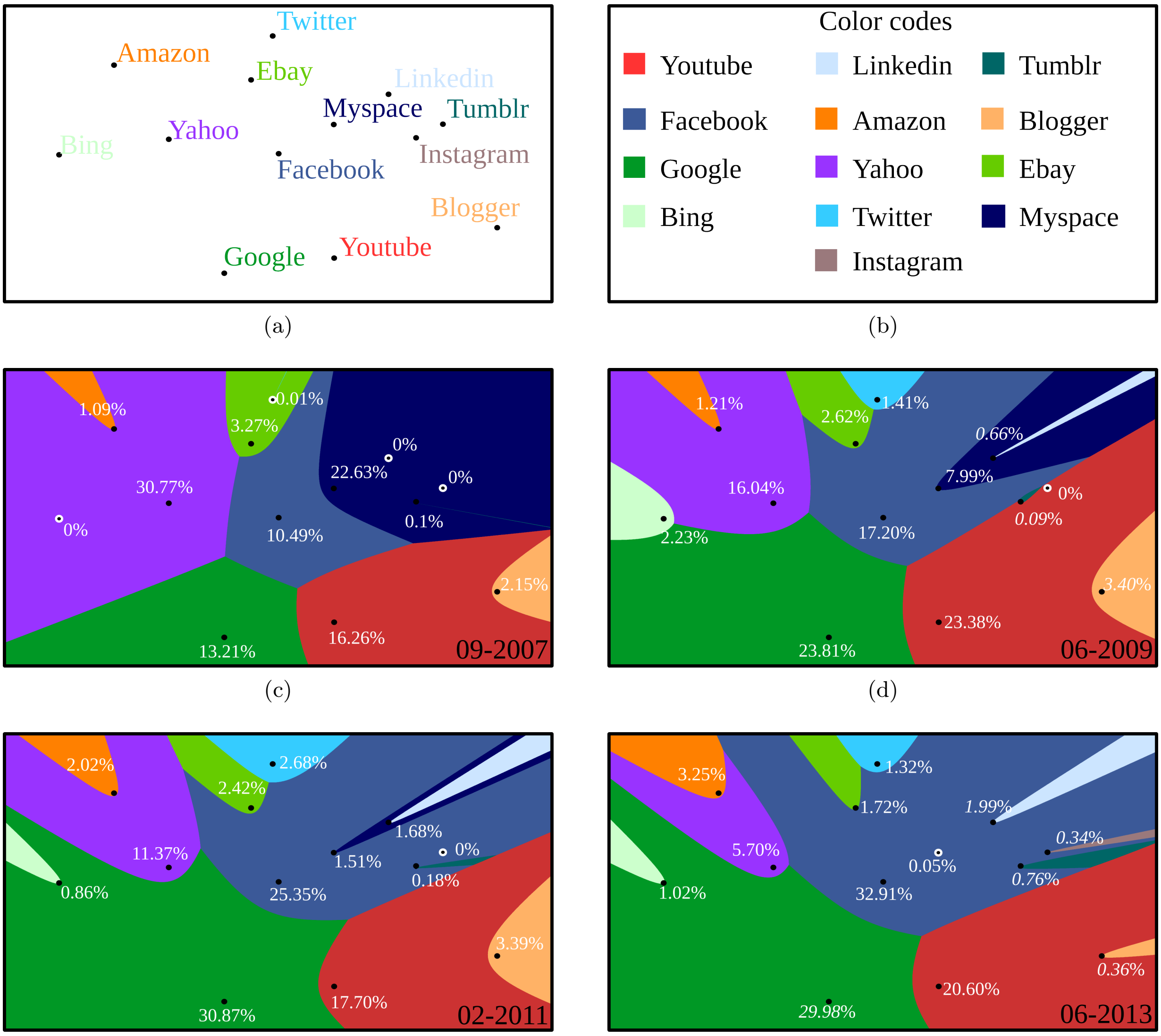
[1] Basu, A., Mitchell, J., and Sabhnani, G., "Geometric algorithms for optimal airspace design and air traffic controller workload balancing," in [*in Proceedings of the 9th SIAM Workshop on Algorithm Engineering and Experiments (ALENEX)*], SIAM (2008).

[2] Chu, H., Wu, Y., Zhang, Q., and Wan, Y., "Colonial algorithm: A quick, controllable and visible one for gerrymandering," in [*Information and Automation*], Qi, L., ed., *Communications in Computer and Information Science* **86**, 424–430, Springer Berlin Heidelberg (2011).

[3] Haugland, D., Ho, S. C., and Laporte, G., "Designing delivery districts for the vehicle routing problem with stochastic demands," *European Journal of Operational Research* **180**(3), 997 – 1010 (2007).

[4] Okabe, A. and Suzuki, A., "Locational optimization problems solved through Voronoi diagrams," *European Journal of Operational Research* **98**(3), 445 – 456 (1997).

[5] Caro, F., Shirabe, T., Guignard, M., and Weintraub, A., "School redistricting: Embedding GIS tools with integer programming," *The Journal of the Operational Research Society* **55**(8), pp. 836–849 (2004).

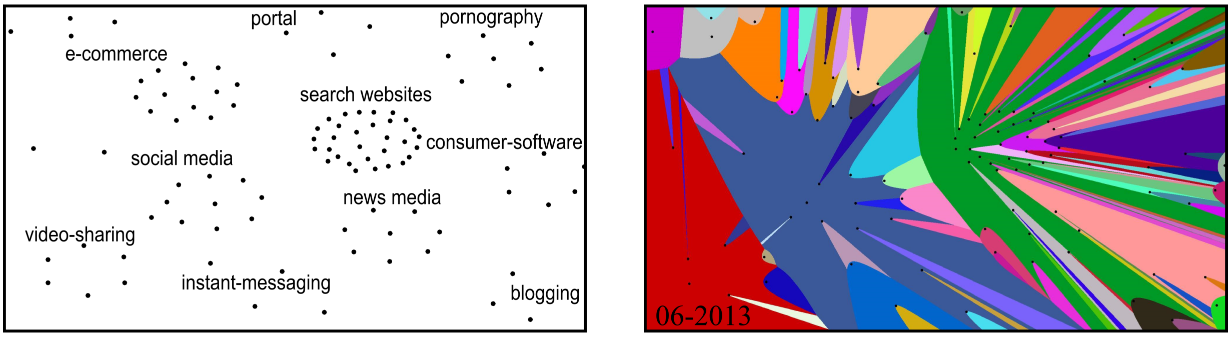


(a) (b)

(c) (d)

(e) (f)

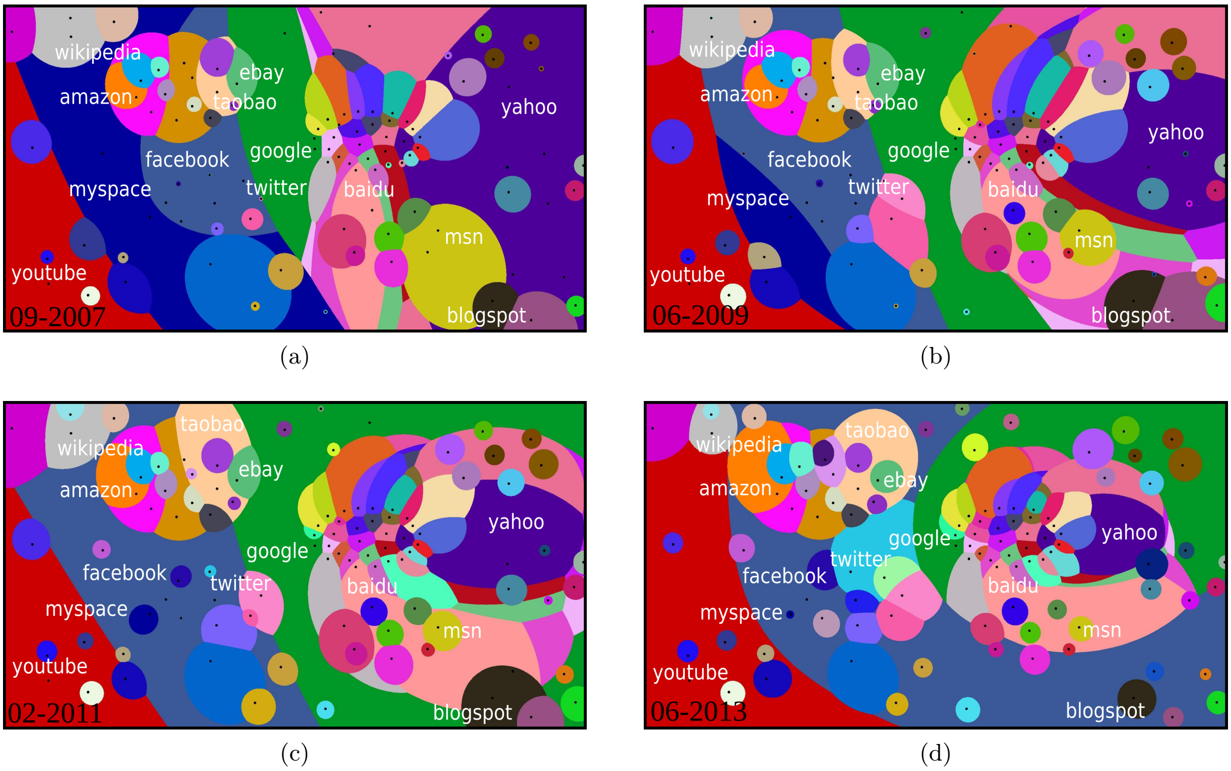
Figure 6: A collection of "maps" that display the relative "sizes" of 13 major websites, taken between 2007 and 2013.



(a) Locations of top 101 websites clustered based on their genre (obtained using *Graphviz*)

(b) A partitioned map of top 101 websites

Figure 7: Locations and additive Voronoi partition for the top 101 internet websites



(a)

(b)

(c)

(d)

Figure 8: A collection of “maps” that display the relative “sizes” of the top 101 major websites, taken between 2007 and 2013

[6] Larson, R. and Odoni, A., [*Urban Operations Research*], Dynamic Ideas (2007).

[7] Andrews, K., Kienreich, W., Sabol, V., Becker, J., Droschl, G., Kappe, F., Granitzer, M., Auer, P., and Tochtermann, K., “The infosky visual explorer: Exploiting hierarchical structure and document similarities,” *Information Visualization* **1**, 166–181 (2002).

[8] Brner, K., Chen, C., and Boyak, K. W., “Visualizing knowledge domains,” *Annual Review of Information Science and Technology* **37**, 179–255 (2003).

[9] Boots, B. and South, R., “Modeling retail trade areas using higher-order multiplicatively weighted voronoi diagrams,” *Journal of Retailing* **73**, 519–536 (1997).

[10] Couclelis, H., “Worlds of information: The geographic metaphor in the visualization of complex information,” *Cartography and Geographic Information Systems* **25**, 209–220 (1998).

[11] Kaski, S., Honkela, T., Lagus, K., and Kohonen, T., “Websom: Self-organizing maps of document collections,” *Neurocomputing* **21**, 101–117 (1998).

[12] Skupin, A., “A cartographic approach to visualizing conference abstracts,” *IEEE Computer Graphics and Application* **22**, 50–58 (2002).

[13] Tobler, W., “Thirty five years of computer cartograms,” *Annals of the Association of American Geographers* **94**, 58–73 (2004).

[14] Chen, H., Houston, A. L., Sewell, R. R., and Schatz, B. R., “Internet browsing and searching; user evaluations of category map and concept space techniques,” *Journal of the American Society for Information Science* **49**, 582–608 (1998).

[15] Reitsma, R., Trubin, S., and Mortensen, E., “Weight-proportional space partitioning using adaptive voronoi diagrams,” *Geoinformatica* **11**(3), 383–405 (2007).

[16] Carlsson, J., Carlsson, E., and Devulapalli, R., “Shadow prices in territory division,” *INFORMS Journal on Computing* **Under revision** (2013). see `http://menet.umn.edu/~jgc/shadow-prices-rev2.pdf`.