Adaptive Multiplicatively Weighted Voronoi Diagrams for Information Space Regionalization

René Reitsma[*]
College of Business
Oregon State University

Stanislav Trubin[†]
Electrical Engineering & Computer Science
Oregon State University

Saurabh Sethia[‡]
Electrical Engineering & Computer Science
Oregon State University

Abstract

In an attempt to regionalize and visualize regions of information space, we offer a method for regionalization based on predetermined area relationships. This problem can be conceptualized as an inverse multiplicatively weighted Voronoi diagram. To compute such a diagram, we offer an adaptive algorithm that minimizes area deviations. We apply the method to a spatialization of a web site for undergraduate engineering learning.

CR Categories: H.1.m [Models and Principles]: Miscellaneous—Information Space; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/Methodology

Keywords: information space, spatialization, Voronoi diagrams, adaptive systems

1 Introduction: Information Space Regionalization

To improve navigation and assessment of information systems, authors from various disciplines are exploring the applicability and usefulness of cartographic techniques for visualizing and mapping abstract information spaces [Gatrell 1983; Couclelis 1998; Chen 19; Dodge and Kitchin 2001; Fabrikant and Buttenfield 2001; Kohonen 2001]. Since these spaces cannot be directly observed – they are largely methodological artifacts – their dimensionality, metric and extent are either a priori chosen [Mukherjea and Hara 1997; Card et al. 1996; Lamping and Rao 1996; Munzner 1998; Kahn and Lenk 2000; Huffaker et al. 2002] or instead inferred from secondary information.

An example of the latter is a method proposed by Buttenfield & Reitsma [2002] and Reitsma et al. [2004] that uses information from the navigational logs of an information system to reconstruct the system's spatial dimensionality and metric. The method is based on the spatial metaphor that the more users navigate between any two items in an information system, the closer the two items are positioned in the associated information space. For example, if a request for Web page *i* is more often followed by a request for Web page *j* than for a request for other pages, *i* and *j* are considered 'close' to each other. From the relative proximities of all Web pages a spatial representation can then be computed. The authors demonstrate the applicability of the approach with empirical study of a digital library of cartographic materials [Buttenfield and Reitsma 2002] and an undergraduate engineering learning information system [Reitsma et al. 2004]. In both cases an interpretable, four-dimensional information space is reconstructed from the system's navigational logs after which the system's information items; e.g., web pages are fitted into this space.

Although the above method can be used to visualize an information system's use-space, this space remains uniform and unmodeled. With this we mean to say that although we can locate the information items in the space, the space itself remains uniform, featureless and contains no zones or regions. Yet, one could consider the space around the information items basins of attraction; i.e., spheres of influence or regions that are characterized by some of the same properties as those of their central items. A similar notion is present in the classic regional geographical studies by Christaller [1933], Isard [1967], Berry [1967] and later by Huff [1967] and Huff & Lutz [1979]. For a summary of some of these studies, refer to Hagget [1972]. Interestingly, a byproduct of the method by Buttenfield and Reitsma [2002] is the estimation of a parameter for each information item that can be interpreted as the relative size or area of its sphere of influence. Hence, in this paper we propose a method for reconstructing such spheres of influence based on these given areas or sizes. To accomplish this, we invert a classic spatial allocation model, namely that of the multiplicatively weighted Voronoi diagram (MWVD) [Okabe et al. 2000; Boots 1980]. We show results indicating that the method works well and then use it to regionalize a spatialization of a web site for undergraduate engineering learning.

2 Voronoi Diagrams

A classic way of regionalizing or allocating space around a predetermined set of points or 'generators' is Voronoi tessellation. This method is based on finding the nearest generator for every point in the space. The resultant regionalization is known as a Voronoi diagram (VD). In a comprehensive presentation on the subject including an extensive review of the literature, Okabe et al. [2000] present the ordinary VD as well many of its generalizations. Table 1 presents a few of the characteristics of both the ordinary and multiplicatively weighted Voronoi diagram.

For our objective of regionalizing an information space based on the location of a set of information items or generators and their predetermined area, the MWVD is intuitively attractive. 'Area' or 'size' is a multiplicative property, we want to allocate all of the available space and we like the basins to grow as uniformly as possible in all directions; i.e., we would like to preserve their circular, disk-like shapes as much as possible. Furthermore, the required computations are simple, especially when carried out in a raster fashion. However, the standard MWVD method has some problems.

[*]e-mail: reitsmar@bus.oregonstate.edu
[†]e-mail: trubin@eecs.oregonstate.edu
[‡]e-mail: saurabh@eecs.oregonstate.edu

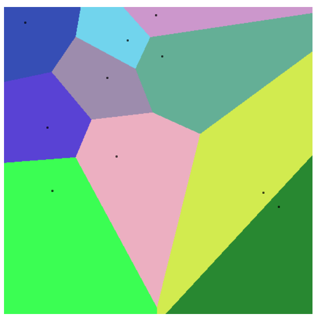
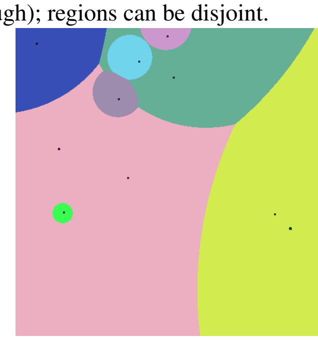
	Definition
A. Ordinary Voronoi Diagram	Formal: $V_i = \{x \mid \|x - x_i\| = \|x - x_j\|$ for $i \neq j$, where x_i and x_j are location vectors and V_i is the Voronoi region for point i $\}$ Dynamic interpretation: All generators simultaneously start growing and grow at identical rates. Characteristics: Borders are straight; regions are contiguous. 
B. Multiplicatively Weighted Voronoi Diagram (MWVD)	Formal: $V_i = \{x \mid \|x - x_i\|/w_i = \|x - x_j\|/w_j$ for $i \neq j$, where x_i and x_j are location vectors and V_i is the Voronoi region for point i $\}$ Dynamic interpretation: All generators simultaneously start growing but they grow at different rates represented by their weights. Characteristics: Borders are curved; regions can surround other regions (if they grow fast enough); regions can be disjoint. 

Table 1: Ordinary and Multiplicatively Weighted Voronoi Diagrams.

3 Inverse MWVDs

Our problem, i.e., to generate Voronoi regions whose relative area sizes are known, is essentially the inverse of the traditional MWVDs. Whereas for a traditional MWVD *area* is the dependent variable, we want to find a set of *growth* rates resulting in given *area* ratios. This changes the problem from prediction to optimization; i.e., compute regions such that their areas optimally represent a predetermined set of values. Since in the dynamic representation of a MWVD the weights represent expansion rates, we can start by adjusting area weights using (1).

$$\sqrt{area/\pi} \tag{1}$$

However, this assumes unhindered, uniform and omnidirectional growth, something which only applies until equilibrium between adjacent generators is reached; i.e., until two expanding generators 'meet.' It also does not account for regional containment or nesting, a phenomenon that occurs when one generator grows so quickly that it envelops a slower growing, neighboring generator (see the example in Table 1B).

An additional problem concerns the areas associated with peripheral generators. Since peripheral generators can grow unchecked in the opposite direction of the center of gravity of the generator configuration as a whole, their areas become infinitely large. This problem, of course, applies to VDs in general and is normally resolved in either of two ways:

- A set of auxiliary generators is regularly spaced beyond the perimeter of the original generators. When the VD gets solved, the regions associated with the auxiliary regions will extend to infinity, but the regions of the originally peripheral generators now have boundaries. This is the approach that must be used in convex hull applications such as *qhull* [Barber et al. 1996].
- Introduce an artificial outer boundary and limit the growth of the external generators to these boundaries. This is the approach used in most GIS systems; e.g., *ArcGIS* or *IDRISI* [DeMers 2000].

In either case, outward growth of the peripheral generators is limited by an *ad hoc* decision of the modeler; either auxiliary generators are positioned at *ad hoc* locations, or an ad hoc boundary is chosen. Regardless of the solution, however, the areas of the resulting regions are unlikely to properly reflect their target values.

Table 2 (MWVD) illustrates the limiting success of the above approach. It shows the average results of computing MWVDs for 100 sets of $N = 25$ points with weights $w_{j+1} = w_j + 1$, $w_0 = 1$ and $j = 0, 1, \ldots, N-1$. To prevent peripheral areas from attaining infinitely large areas, a rectangular boundary was selected. Positional information of the points was selected randomly from a uniform distribution with a minimum of zero and a maximum equal to the maximum coordinates of the bounding rectangle. Table rows show four types of error: correlation between weight and region area (r) and minimum, average and maximum error where error is defined as in (2).

$$Error = |A_j - a_j|/A_j \tag{2}$$

where:

A_j = target area of generator j as proportion of total area.

a_j = accomplished area of generator j as proportion of total area.

	MWVD		AMWVD	
	area	$\sqrt{area}$	*area*	$\sqrt{area}$
Iterations			1000	1000
r	0.767	0.667	0.954*	0.948*
Min. Error	0.045	0.032	< 0.001*	< 0.001*
Mean Error	0.585	0.500	0.037*	0.039*
Max. Error	1.490	1.868	0.331*	0.369*
* paired two-tailed t-test with $p < 0.001$				

Table 2: Performance of MWVD and AMWVD data.

Note that for the MWVD the correlations do not reach above .77 and that the adjustment of (1) does not uniformly improve matters.

4 Adaptive Multiplicatively Weighted Voronoi Diagrams (AMWVD)

To better approach predetermined area relationships we propose an iterative, adaptive solution, where the MWVD is solved repeatedly and for each new solution the weight set is adapted based on the error of the previous solution. This is represented by (3).

$$w_{i+1,j} = w_{i,j} + \Delta w_j \tag{3}$$

where $w_{i,j}$ is the weight of generator j at iteration i.

Δw_j is positive if generator j received less area than represented by $w_{0,j}$; Δw_j is negative if too much area was allocated. Weights at iteration $i = 0$ represent the area targets; $A_j = w_{0,j}$. We define $a_{i,j}$ as the area allocated to generator j after iteration i. Both A_j and $a_{i,j}$ are proportions and are constrained by (4) and (5) respectively.

$$A_j \in (0,1), \text{and} \sum_{j=0}^{N-1} A_j = 1 \tag{4}$$

where N is the number of generators.

$$a_{i,j} \in (0,1), \text{and} \sum_{j=0}^{N-1} a_{i,j} = 1 \tag{5}$$

where $i = 0, 1, \ldots, I-1$.

Since the notion of a AMWVD is to continuously adjust weights such that the difference $A_j - a_{i,j}$ gets minimized, we suggest making Δw_j proportional to the following three quantities:

1. $w_{i,j}$,
2. the difference in target and allocated areas $A_j - a_{i,j}$, and
3. a positive, nonzero scaling constant k.

After implementing these recommendations, we obtain the following expression for adaptive weights (6):

$$w_{i+1,j} = w_{i,j} + w_{i,j} k(A_j - a_{i,j}) \tag{6}$$

After regrouping of terms (7):

$$w_{i+1,j} = w_{i,j}(1 + k(A_j - a_{i,j})) \tag{7}$$

5 Choosing k

We offer the following heuristic for selecting k. The worst case obtains when at some iteration i a region j occupies almost all of the available space while having a low weight, or when a region j occupies very little space while having a high weight. Thus, the difference $A_j - a_{i,j}$ lies between negative one and one (8):

$$-1 < A_j - a_{i,j} < 1 \tag{8}$$

After multiplying by k and adding 1, we obtain (9):

$$1 - k < A_j - a_{i,j} < 1 + k \tag{9}$$

Accordingly, choosing $k = 1$ is a safe default as it will not result in negative weights. If we take $k = 1$, (7) reduces to (10):

$$w_{i+1,j} = w_{i,j}(1 + A_j - a_{i,j}) \tag{10}$$

However, it is important to note that for some data sets k can be chosen larger than 1.0. In these cases the difference $A_j - a_{i,j}$ may get minimized in fewer iterations and the adaptive scheme converges more quickly. We have also found that using a different value for k sometimes prevents the algorithm from getting stuck on what appears like a local minimum. We are in the process of investigating these behaviors.

6 Results

Table 2 (AMWVD) shows the results of using the adaptive method for the same 100 generator point configurations that were used to compute the ordinary MWVDs (Table 2: MWVD). Runs of 1000 iterations each were made for starting weights $w_{0,j} = A_j$ and $w_{0,j} = \sqrt{A_j}$. k was set to 1.0 for all runs. Comparing the values in columns MWVD and AMWVD, all error measures using the adaptive method are significantly better than those using the regular MWVDs ($p < 0.001$).

Figure 1 shows data indicating how an AMWVD solution for a random set of 25 points converges by displaying the trajectory of the error measures used in Table 2 and computed using (2) and the adaptive scheme of (10). The graph shows that the correlation between target area and allocated area (*rAa*) quickly approaches 1.0. However, the mean error takes quite a bit longer to approach 0.0. The worst fitting region takes yet longer to converge.

7 Case Study: The Building as a Learning Tool (BLT)

We applied the above method on a spatialization of a web site for undergraduate engineering learning maintained at the University of Colorado's College of Engineering and Applied Science [Carlson et al. 2003; Reitsma et al. 2004]. The site, http://blt.colorado.edu provides real-time, web-based access to a variety of building control data that students and instructors use to study engineering principles and methods. Figure 2 shows the result of a query submitted to the system; the last 24 hours of temperature data collected on one of the building walls.

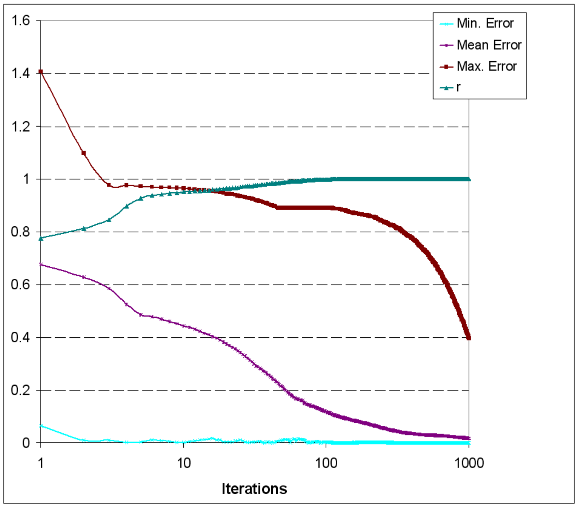


Figure 1: Error trajectories for AMWVD.

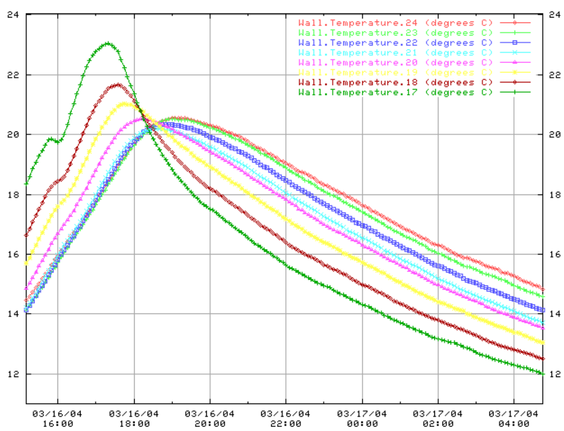


Figure 2: A sample query from the BLT system.

Using the techniques proposed by Buttenfield and Reitsma [2002], Reitsma et al. [2004] derived a four-dimensional spatialization of the web-site:

Dimension 1: Size. The first dimension represents the size of web pages expressed as the sheer volume of requests they receive.

Dimension 2: Active vs. Passive. The second dimension represents active vs. passive types of user interactions. BLT pages that do not allow for much user interaction have dimensional scores opposite of those that do.

Dimension 3: Information Generalization and Density (High vs. Low detail). This dimension represents the depth or level of detail offered by a BLT web page. Pages at one end of the dimension provide a general overview of the BLT information and offer very little detail. The other end of the dimension contains pages with lots of specific, high density information.

Dimension 4: Information Type (Sensor/System Data). The fourth dimension scales the nature of the data; i.e., sensor data vs. system data. Web pages describing the BLT system are located at one end of the dimension, whereas pages containing sensor information are clustered at the other end.

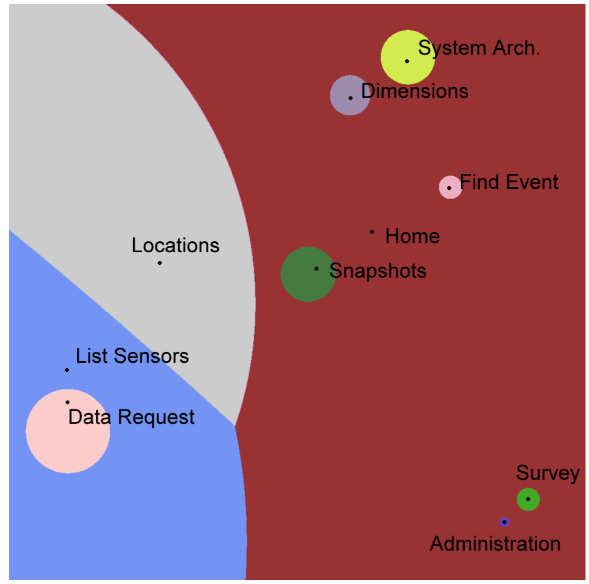


Figure 3: MWVD of BLT spatialization (dimension 1 vs dimension 2).

Page Group	τ
Home	4.776
System Architecture	.727
Locations	3.602
Dimensions	.697
List Sensors	3.552
Find Event	0.616
Survey	0.178
Administration	0.072
Snapshots	1.721
Data Request	2.441

Table 3: BLT Page groups and their request rates.

8 Regionalization of the BLT space

Figure 3 and figure 4 show the MWVD and AMWVD regionalizations of the BLT information space. Dimension 1 represents the size of the Web page groups as measured by the amount of traffic they attract. Dimension 2 differentiates pages that require user input (active pages) from those that do not (passive pages). The coordinates for the page groups were derived by means of a multi-dimensional scaling following the methods proposed by Buttenfield and Reitsma [2002]. As mentioned earlier, this method also yields a parameter τ for each web page or group of pages representing the volume of requests it receives (Table 3). This parameter was used for w_i in the MWVD (figure 3) and A_j in the AMWVD (figure 4). The AMWVD was computed using 1000 iterations and $k = 0.5$. Note how the area distribution of the AMWVD in figure 3 is quite different from that of the nonadaptive MWVD in figure 4.

Figure 5 shows the trajectory of the goodness-of-fit measures over 1000 adaptations under the AMWVD method.

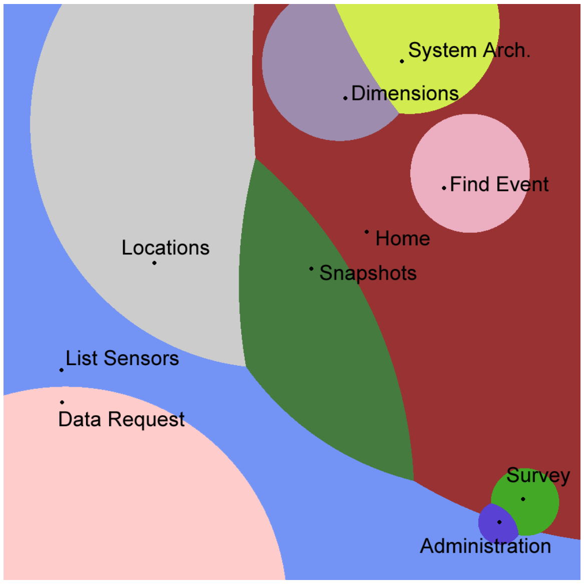


Figure 4: AMWVD of BLT spatialization (dimension 1 vs dimension 2).

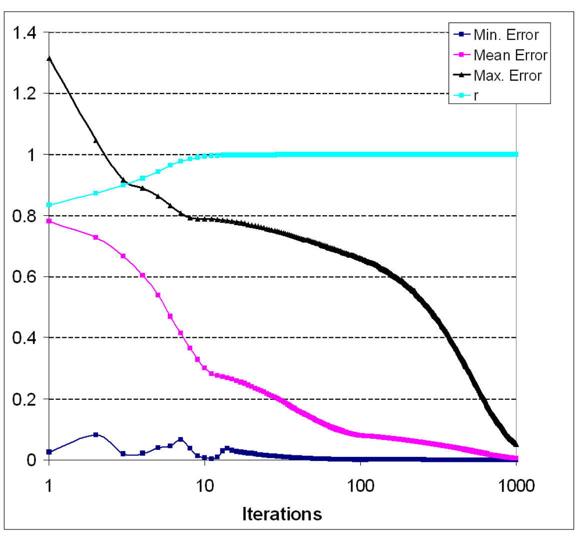


Figure 5: Error trajectories for AMWVD of BLT data.

9 Conclusions

The methods described by Buttenfield and Reitsma [2002] and Reitsma et al. [2004] provide a measure of the size or area of basins of attraction in information space. In reconstructing these basins MWVDs offer a good starting point because their multiplicative nature reflects the nature of the size/area measures and because it allows generators to 'grab' space beyond neighboring generators while preserving the principle of uniform growth in all directions. We overcame the main disadvantage of the MWVD; i.e., weights reflecting growth rate rather than resultant area, by using an adaptive, iterative approach, adjusting the weight set based on the error of the previous solution.

Both Monte Carlo testing and the modeling of the BLT web site spatialization yield good results. The goodness-of-fit measures quickly approach their desired values and the solutions are relatively stable.

10 Acknowledgements

This work was funded by the Integrated Teaching and Learning Laboratory of the University of Colorado, Boulder and the US Dept. of Education's Fund for the Improvement of Post Secondary Education (FIPSE).

References

BARBER, C., DOBKIN, D., AND HUHDANPAA, H. 1996. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software 22*, 469–484.

BERRY, B., AND LOBLEY, J. 1967. *Geography of Market Centers and Retail Distribution*. Prentice Hall.

BOOTS, B. 1980. Weighting thiessen polygons. *Economic Geography 56*, 248–257.

BUTTENFIELD, B., AND REITSMA, R. 2002. Loglinear and multidimensional scaling models of internet transactions. *International Journal of Human-Computer Studies 57*, 101–119.

CARD, S., ROBERTSON, G., AND YORK, W. 1996. The webbook and the webforager; an information workspace for the world-wide web. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 96)*, ACM Press.

CARLSON, L., REITSMA, R., BRANDEMUEHL, M., HERTZBERG, J., SULLIVAN, J., AND GABBARD, S. 2003. Exploiting an engineering building as a unique distance learning tool. *International Journal of Engineering Education 19*, 379–388.

CHEN, C. 19. *Information Visualization and Virtual Environments*. Springer, London, UK.

CHRISTALLER, W. 1933. *Die zentralen Orte in Suddeutschland*. Gustav Fischer, Jena, Germany.

COUCLELIS, H. 1998. Worlds of information: the geographic metaphor in the visualization of complex information. *Cartography and Geographic Information Systems 25*, 209–220.

DEMERS, M. 2000. *Fundamentals of Geographic Information Systems*. Wiley & Sons, NY, NY.

DODGE, M., AND KITCHIN, R. 2001. *Mapping Cyberspace*. Routledge, London, UK.

FABRIKANT, S., AND BUTTENFIELD, B. 2001. Formalizing semantic spaces for information access. *Annals of the Association of American Geographers 91*, 263–280.

GATRELL, A. 1983. *Distance and Space; a Geographical Perspective*. Clarendon Press, Oxford, UK.

HAGGET, P. 1972. *Geography; a Modern Synthesis*. Harper & Row, NY, NY.

HUFF, D., AND LUTZ, J. 1979. Ireland's urban system. *Economic Geography 55*, 196–212.

HUFF, D. 1967. The delineation of a national system of planning regions on a basis of urban spheres of influence. *Regional Studies 7*, 323–329.

HUFFAKER, B., FOMENKOV, M., PLUMMER, D., MOORE, D., AND CLAFFY, K. 2002. Distance metrics in the internet. In *IEEE International Telecommunications Symposium*, IEEE.

ISARD, W. 1967. *Models of Regional Analysis. Introduction to Regional Science*. MIT Press, Cambridge, MA.

KAHN, P., AND LENK, K. 2000. *Mapping Web Sites: Digital Media Design*. Rotovision.

KOHONEN, T. 2001. *Self-Organizing Maps*. Springer, Berlin, Germany.

LAMPING, J., AND RAO, R. 1996. Vizualizing large trees using the hyperbolic browser. In *Proceedings ACM CHI'96*, ACM Press, NY, NY.

MUKHERJEA, S., AND HARA, Y. 1997. Focus + context views of world-wide web nodes. In *Proceedings of Hypertext 97*, ACM Press, 187–196.

MUNZNER, T. 1998. Exploring large graphs in 3d hyperbolic space. *IEEE Computer Graphics and Applications 18*, 18–23.

OKABE, A., BOOTS, B., SUGIHARA, K., AND CHIU, S. 2000. *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*. Wiley & Sons, Chichester, UK.

REITSMA, R., THABANE, L., AND MACLEOD, M. 2004. Spatialization of web sites using a weighted frequency model of navigation data. *Journal of the American Society for Information Systems and Technology 55*, 13–22.