Adapted Dorling cartogram on wage inequality in Portugal

Pedro Cruz

College of Arts, Media and Design, Northeastern University

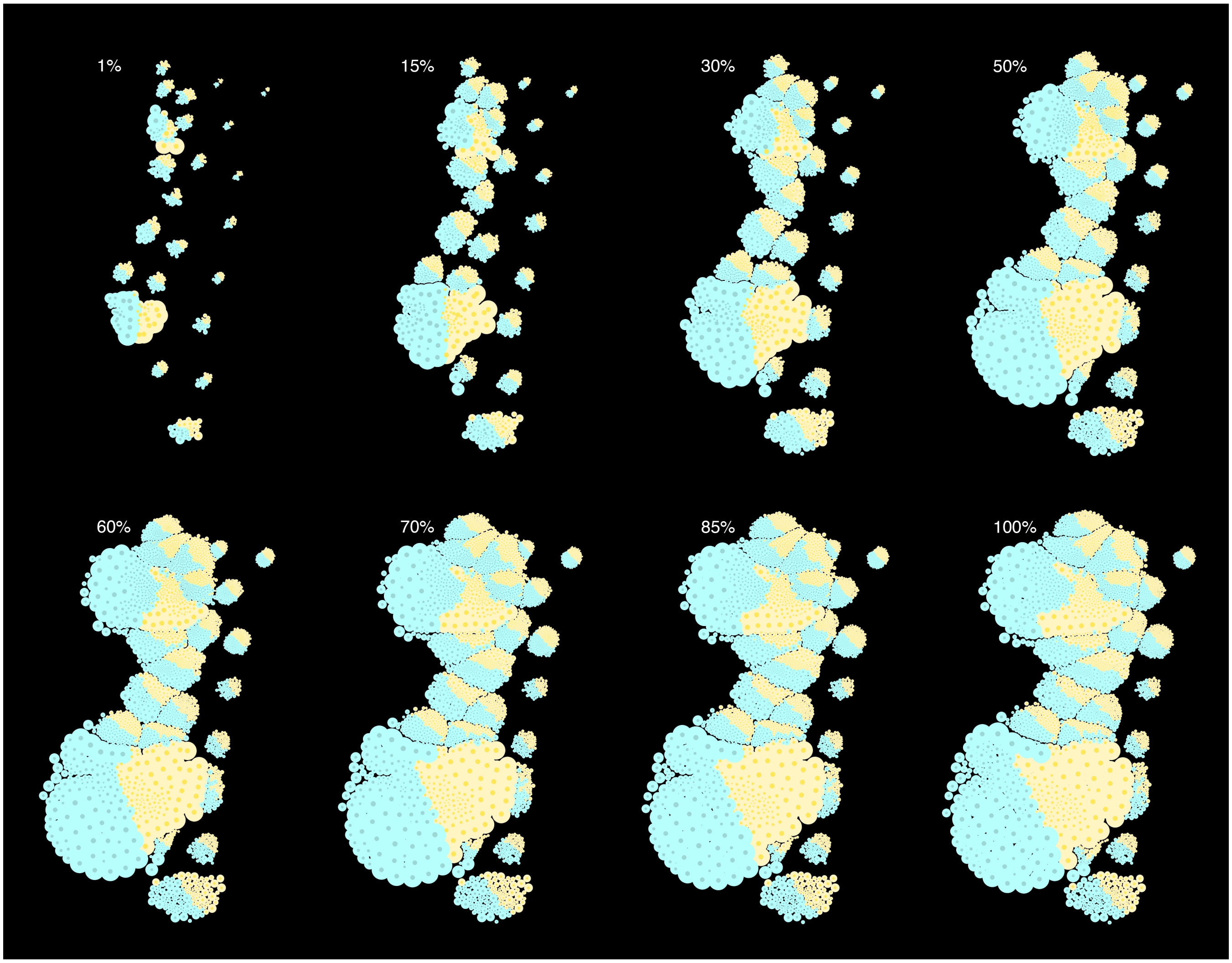

Figure 1: Several stages in the formation of a cartogram on wage inequality in Portugal (2003) as parishes are inserted and constraints are met. After 60% of the simulation's duration, all the parishes are present. Total wages earned by men in blue, by women in yellow.

Abstract

Wage inequality in Portugal has been systematically around 27%. This paper describes a set of animated cartograms that show this inequality for several years. The visualization was designed for an exhibition setting, and instead of presenting the final cartograms, it shows the cartogram formation for each year as a way of reiterating the message. The model used is a modification of Dorling cartograms that visually appears as a set of contiguous amalgams of total wages earned by subregion. This aspect confers an organic tone to the artifact, as if liquids of different densities were being poured into the canvas without mixing, and depicting a tension-based behavior where amalgams of men and women push among themselves to compete for the same space.

* p.cruz@northeastern.edu

[1] www.londondesignbiennale.com

1 Context

The visualization described here was part of the Portuguese exhibition in the London Design Biennale 2016[1]. The exhibition title was "UN/BIASED", and explores the biennale's theme "utopia by design" by presenting utopian views together with contrasting problematic realities in Portuguese society. One of such realities is wage inequality between genders: The focus of the visualization presented herein.

Recent statistics report wage inequality around 27%, without much change throughout several years. Cartograms distort maps to show statistical variables, and by doing so, they distort an image of a familiar representation embedded in our culture [1]. This visualization shows the reality of wage inequality in Portugal as a distortion of what would be ideal, and for such, it uses a cartogram. The cartogram partitions space into separate amalgams of wages by gender that vary in size, but that are confined in the same space. The objective of the partitioning and confinement is to portray the hindrance of equality through the unequal distribution of wealth between both genders. This cartogram on wage inequality was implemented using a modified version of the Dorling cartogram model [1] in order to better enable the comparison between genders while conferring organic aesthetics to the artifact.

The piece does not present a final cartogram for quick observation, but instead shows how the cartogram is built in order to provoke curiosity on its final form and meaning (see figure 1). The visualization was designed for an exhibition setting and a scenario where the viewer is presented with a specific form, that because of its unfamiliarity, and its organic aesthetics, can be seen as intriguing and call attention to the matter. Although the piece aims at creating awareness on wage inequality in Portugal, it is not of a utilitarian nature, instead existing for the viewers' contemplation. From this perspective, this visualization fits in the umbrella of "casual visualization" [4].

2 Visualization

The piece shows a series of four animated cartograms consecutively for 2009, 2010, 2011, 2012, and 2013. This piece does not interpolate cartograms from 2009 through 2013 as a way to show data variations, but instead shows the computational process of the cartogram formation for each year, resetting when presenting the follow-up year. These cartograms result in visualizations that in form are between contiguous value-by-area cartograms [2], and spatially discrete Dorling cartograms [1].

2.1 Information

The dataset refers to average monthly wages earned by non-freelance workers, per parish[2] by gender. The data was collected from PORDATA[3], for the years 2009-2013. In Portugal, on a national level, men make on average about 27% more than women as can be seen in table 1. This difference has remained constant throughout the collected years. As such, the visualization described here does not aim at showing the differences between years, but instead reiterates the same situation year after year.

Table 1: Monthly Average Wages Earned by Men and Women From 2009 to 2013.

	Men	Women	Gap
2009	€1139	€899	26.6%
2010	€1185	€937	26.5%
2011	€1195	€946	26.4%
2012	€1212	€956	26.8%
2013	€1209	€958	26.2%

Mainland Portugal is organized in 23 statistical subregions. A detailed vector map of these Portuguese regions was simplified[4] by using a path simplification algorithm that reduces the number of vertices in the paths, resulting in the simpler outlines seen in figure 2. Each subregion has a certain number of parishes, totaling 2,882 for the whole country. The parish is the finest aggregation level in the dataset. Each parish has men workers, and women workers. Men workers have an average wage, while women workers have a different average wage. Multiplying the corresponding average wages by the number of workers, gives the total paid in wages. This total by gender, together with the workforces' size, is the information that makes the cartogram.

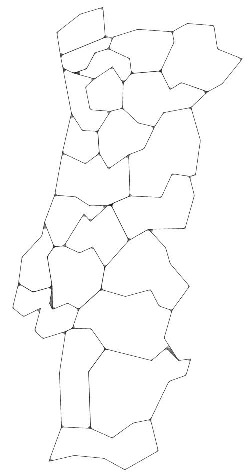

Figure 2: The 23 statistical subregions in mainland Portugal, after applying a path simplification algorithm.

2.2 Representation

Each parish has two corresponding circles: one with size proportional to the total wages of men (in light cyan) and another with size proportional to the total wages of women (in light yellow). Inside each of these circles, stays a smaller one that represents the number of men or women workers (figure 3). This is included in the visualization to show that the number of working men and women is similar, in spite of the discrepancies of total wages earned by gender.

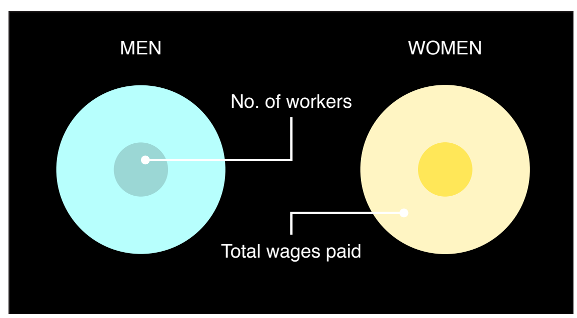

Figure 3: Each parish is represented by two circles, one for men in light cyan, and the other for women in light yellow. The area of each circle represents the total wages earned by each gender. The smaller circle superimposed represents the women or men workers for the corresponding parish.

2.3 Simulation

As the simulation runs for each year, for each frame, the system goes through each subregion and inserts one pair of circles corresponding to a parish of that subregion. This means that in one frame several parishes can be inserted in the simulation at the same time. The pair stays in the canvas for the rest of the simula-

[2] These territorial divisions are officially called "freguesias". Although historically "freguesias" had their origins in the religious organization of the country into parishes, nowadays they do not necessarily correspond one to another. The term parish is used herein because it conveys a granularity level that can partition a mid-sized city into several units.

[3] www.pordata.pt.

[4] http://mapshaper.org.

tion. The system inserts only parishes that have not been inserted before. Since each subregion has a different number of parishes, some subregions may be fully represented in the canvas before others. After 1,064 frames, all the 2,882 parishes have been inserted.

The circles are placed near the centroid of the corresponding subregion, which was previously computed. If the circle represents men, it is placed southwest of the centroid, and if it represents women, it is placed northeast of the centroid. These positioning points remain fixed throughout each simulation run, but are randomized by 2 pixels in each direction each time the simulation is started, which causes variations in the output later discussed in this paper. This small randomization of 2 pixels ensures that the de-overlapping algorithm discussed later on pushes the circles away from each other. Other smaller non-zero randomizations would also work, while randomizations as big as the radiuses of the circles would compromise the appearance of the visualization – it would no longer appear as contiguous areas, but instead as a series of disconnected circles. With this, circles are positioned in distinct regions, creating amalgams of either men or women that can be distinguished and compared. If the circles were to be positioned in the geographical locations of the corresponding parish, this would impede the direct observation of the 27% wage disparity, and prevent the portrayal of one group occupying more space than the other.

The circles in the visualization canvas, do not have mass, velocity, nor acceleration, and thus their positions are not simulated though a numerical integration method used in physics simulations. Instead, the positioning of the circles is directly altered in a way that reflects a series of constraints, which in turn result in the visual rendering of a system with apparent physical properties. One of such constraints is the avoidance of circle overlapping. A simple way to this is: If two circles are intersecting in a certain overlapping length, they are pushed in diametric opposing directions by 1/2 of their overlapping length, as illustrated in figure 4. Naturally, forcing any pair of circles to do this, implies a repositioning that may in turn result in other overlaps with circles in the vicinity of the pair. These new overlaps need to be solved as well, are done so by a method called *relaxation* [3], [5] where this constrained repositioning is run for all pairs of circles in the system several times per frame. The higher the number of runs, the most accurate will be the result. In the system, the relaxation loop runs 10 times per simulated frame.

In this visualization, the de-overlapping algorithm does not push the circles by 1/2 of their overlapping length as it is usually done, but instead uses an empirically determined factor of 1/200 of the overlapping distance to push each circle in a pair. Using the factor of 1/2 is quicker to solve the overlapping problem, but the 1/200 was purposely chosen to slow down this process. A slower process enables a lengthier temporal overlap of circles, and thus conveys the circles in the canvas in cohesive areas, and not as a set of distinguishable and discrete circles. Since the circles are always inserted at the same points near their corresponding subregion's centroid, the circles that were already there are pushed outward from the centroids, resulting in an overall behavior that resembles the expansion of a liquid mass being poured on the same point.

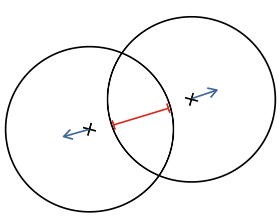

Figure 4: Two overlapping circles and its overlapping length in red. The circles are then pushed by half of this length as shown through the blue vectors. If D is the distance between the center of two circles, with radius R_1 and R_2, then their overlapping distance is simply $(R_1 + R_2)$ - D.

A second constraint in the system is that each amalgam of circles is confined to its corresponding subregion. The boundary of a region acts as an elastic band that expands when geographical space is not enough to accommodate a certain number of circles, but also forcing the circles of a subregion into cohesive amalgams, while trying to prevent the spilling of circles from a region to another. In order to do this, particles of a fixed radius were created for each vertex in the simplified geographic vector map of the subregions, resulting in 2,861 particles. The boundary particles are invisible, but interact with the others created inside the subregions. When a boundary particle and a subregion particle overlap, they are pushed away in the same way as described previously, but this time using the 1/2 length factor to reinforce the boundary's confinement role, making it more resistant to the expansion of a large mass of particles inside a subregion. This non-overlapping constraint for boundary-subregion particles is applied 10 times per frame, similarly as among subregion particles.

The confinement of particles to a subregion cannot be attained solely through the creation of particles over the boundary. In order to turn the boundary to an elastic band, the particles in the boundary need to be connected through strings among themselves. A spring is an elastic device, meaning that when compressed or stretched in relation to its rest length, it exerts a force proportional to its change in length. This coefficient of proportionality is a characteristic of the spring and can be interpreted as its strength. For each spring that connects a pair of particles *a* and *b*, the constraint applied is described in (1), where *a* and *b* are the particles' position vectors, and vectors are in red. This constraint pushes or pulls the particles proportionally to the spring's change in length and proportionally to its strength. The strength used in this system is 0.01. When simulating this system of many interconnected springs, relaxation is used, solving the constraint for each spring in the system, 10 times per frame.

```
ConstraintParticles(spring):
   delta = b - a
   dist = magnitude(delta)
   diff = 1 - spring.restLength / dist                    (1)
   shift = delta * 0.5 * diff * spring.strength
   a = a - shift
   b = b + shift
```

The springs were created by connecting any two particles in the subregions' boundaries that are closer than a certain distance (given the scale of the map of 1980 pixels of height, 150 pixels was used as a threshold distance to connect two particles). This strategy creates a mesh of particles that connects nearby points, making the boundaries more resistant, while contributing to preserve the boundary's shape (figure 5).

The spring and overlap constraints are applied to the circles as they are inserted in the visualization canvas. After 1,064 frames, all circles have been inserted, and the simulation runs until 1,800 frames in order to enable the amalgams of circles to expand, together with the subregions' boundaries that enclose them. Figure 6 shows how the subregions adapt as circles are inserted in different stages of the simulation. Six snapshots of the final animation can be seen in figure 1 for the year 2013, rendering invisible the subregion's boundaries.

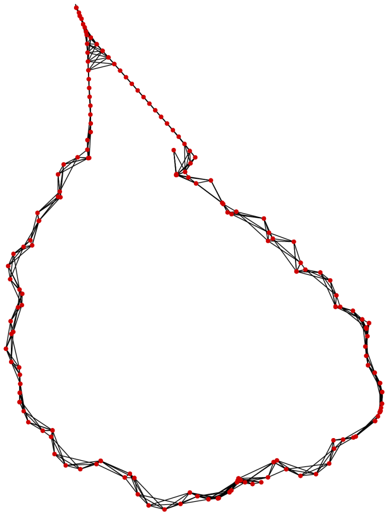

Figure 5: A subregion with its boundary particles interconnected through springs. Two particles are connected if bellow a distance threshold of 150 pixels. This strategy resulted in the creation of 11,494 unique springs, out of the 2,861 particles, that make all the boundaries.

3 Synthesis

This visualization piece shows the distribution of total wealth earned in wages for men and women. For that, it shows wages represented as circles by parish, but confined inside a statistical subregion. The animation shows these circles being inserted into the canvas, while their constraints are computationally applied. The constraints confine the circles inside their corresponding subregion while avoiding spatial overlaps, and with that, the occlusion of circles. It can be observed in figures 1 and 6 that some regions have higher total wages than others due to those areas having larger populations. From this perspective, the cartograms produced here are similar to population cartograms for Portugal. Notwithstanding, it is possible to observe the partition of wages between men and women for each subregion. It should be noticed that in some smaller subregions, where the total of wages distributed is lower, the difference between areas covered by each gender is imperceptible, suggesting wage equality in those cases. However, it should be noted too that there is no observable region where women receive the biggest share of wages.

As the circles are inserted, the subregions formed by elastic bands expand in order to accommodate all the parishes. From a visual perspective, the cartogram being formed is between two visual appearances: the one of a contiguous value-by-area cartogram, and the one of a discrete Dorling cartogram. From a technical perspective, it can be argued that the method is fundamentally a Dorling cartogram, representing statistical areas with non-overlapping circles, but with several modifications. The first is that that for the basic statistical unit, the parish, there are two circles – one for men and another for women. The second modification is that these circles are not initially positioned given the specific geographical location of the corresponding parish. Lastly, the circles are confined to a higher level geographical and statistical subregion, that have boundaries with elastic properties, expanding in order to accommodate all the circles, but also constraining them into cohesive amalgams. These modifications are what makes this cartogram more contiguous at a subregion level than an unmodified Dorling cartogram.

Another important property of this cartogram, is how men and women's circles are positioned differently. As mentioned before, while men are positioned southwest of the centroid, women are positioned northeast of it. This detail makes the system create two distinguishable masses, that enable the comparison of wage distribution between men and women per subregion. These insertion points undergo a slight randomization of 2 pixels at the beginning of each simulation run, conferring to the system generative prop-

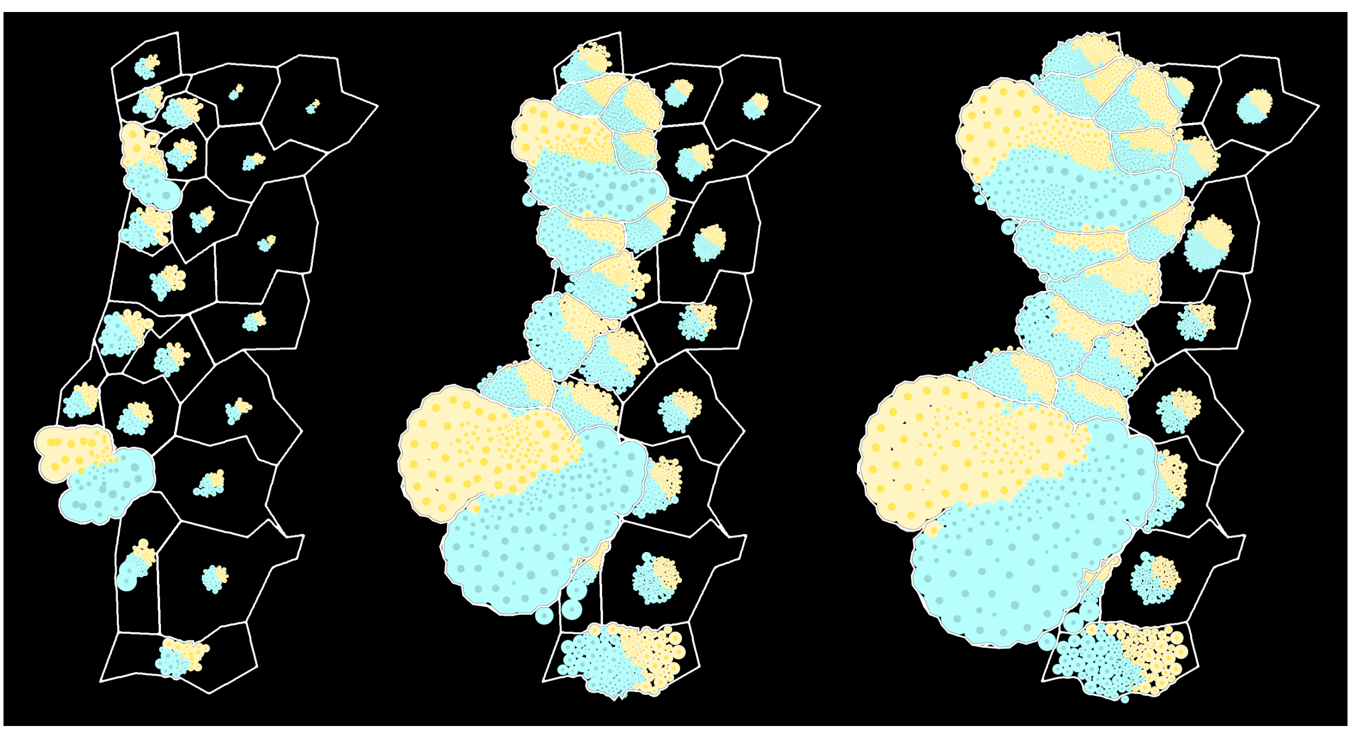

Figure 6: The boundaries of the subregions expanding as circles are inserted in the system (2013 data).

erties: As can be observed between figures 1 and 6, pertaining to the same data of 2013, and configuring different simulation runs, the amalgams do not have the same shape or relative positioning. This variation is caused by the small randomization factor, resulting in clearly different ways to portray the same data while carrying the same meaning.

The five rendered artifacts for years 2009-2013 can be watched in sequence at https://vimeo.com/218215138. Each year is a simulation run of the cartogram. Showing one year after another, with a hard-cut in between, reiterates the same situation year after year. Since the wage gap keeps roughly unaltered during this time period, a system that interpolates through the years would remain largely unaltered, loosing the strong assertion effect that reset and repetition carries. Another alternative that would keep this assertion would be displaying the five simulations simultaneously, but enabling them to quickly loop.

4 Final remarks

This piece was designed to raise awareness for a 27% wage gap in Portugal, but doing so with an emphasis on aesthetics that can provoke attention. The growing amalgams result from the self-organization of circles, that having enough granularity, confer both detail and organized structure to the resulting artifact. The resulting animation resembles two distinct fluids that are poured into the canvas without mixing, as if they had some of fluids' physical properties. This creates a tension-based behavior where the amalgams push among themselves to compete for the same space, where space is here the total amount paid in wages.

The total earned wages by gender is depicted through the areas of irregular shapes. This makes the process of comparison among areas more challenging in order to accurately convey the disparity of 27%. This complexity is intentional, and works as a way for the viewers to discover the differences for themselves. Since this matter is often relinquished as non-significant in so many societal settings, inviting the viewer to spot this difference is also a way to invite them to reflection, and to create their own narrative about how disproportionate they see this distribution. This reflection is steered by the presentation of a figurative image of an anomalous and distorted country.

Acknowledgements

The UN/BIASED exhibition counted with other pieces from artist Marta de Menezes, and was curated by Manuel Lima. The exhibition was produced by Cultivamos Cultura.

References

[1] D. Dorling, "Area cartograms: their use and creation," *Concepts and Techniques in Modern Geography*, vol. 59, 1996.

[2] M. T. Gastner and M. E. J. Newman, "Diffusion-based method for producing density-equalizing maps," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 20, pp. 7499–7504, 2004.

[3] T. Jakobsen, "Advanced character physics," in *Game Developers Conference*, San Francisco, 2001, pp. 383–401.

[4] Z. Pousman, J. T. Stasko, and M. Mateas, "Casual Information Visualization: Depictions of Data in Everyday Life.," *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1145–1152, 2007.

[5] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, 3rd ed. Cambridge, UK: Cambridge University Press, 2007.