ARTICLE IN PRESS
JID: EOR [m5G;May 24, 2017;19:21]

European Journal of Operational Research 000 (2017) 1–17

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

Constrained clustering via diagrams: A unified theory and its application to electoral district design

Andreas Brieden[b], Peter Gritzmann[a,*], Fabian Klemm[a]

[a] Department of Mathematics, Technical University of Munich, München 80290, Germany
[b] Universität der Bundeswehr, Neubiberg 85579, Germany

ARTICLE INFO

Article history:
Received 12 August 2016
Accepted 6 April 2017
Available online xxx

Keywords:
Combinatorial optimization
Constrained clustering
Generalized Voronoi diagrams
Electoral district design
OR in government

ABSTRACT

The paper develops a general framework for constrained clustering which is based on the close connection of geometric clustering and diagrams. Various new structural and algorithmic results are proved (and known results generalized and unified) which show that the approach is computationally efficient and flexible enough to pursue various conflicting demands.

The strength of the model is also demonstrated practically on real-world instances of the electoral district design problem where municipalities of a state have to be grouped into districts of nearly equal population while obeying certain politically motivated requirements.

1. Introduction

Constrained clustering. General clustering has long been known as a fundamental part of combinatorial optimization and data analytics. For many applications (like electoral district design) it is, however, essential to observe additional constraints, particularly on the cluster sizes. Accordingly, the focus of the present paper is on *constrained clustering* where a given weighted point set X in some space $\mathcal{X}$ has to be partitioned into a given number k of clusters of (approximately) *predefined weight*.

As has been observed for several applications, good clusterings are closely related to various generalizations of Voronoi diagrams; see e.g. Brieden and Gritzmann (2012), Geiß, Klein, Penninger, and Rote (2014), Carlsson, Carlsson, and Devulapalli (2016) for recent work that is most closely related to the present paper. Besides electoral district design, such applications include grain-reconstruction in material sciences (Alpers, Brieden, Gritzmann, Lyckegaard, & Poulsen, 2015), farmland consolidation (Borgwardt, Brieden, & Gritzmann, 2011; 2014; Brieden & Gritzmann, 2012), facility and service districting (Aronov, Carmi, & Katz, 2009; Galvão, Novaes, Souza de Cursi, & Souza, 2006; Kalcsics, Melo, Nickel, & Gündra, 2002; Mulvey & Beck, 1984; Muyldermans, Cattrysse, Van Oudheusden, & Lotan, 2002; Segal & Weinberger, 1977; Zoltners & Sinha, 1983), and robot network design (Carlsson, Carlsson, & Devulapalli, 2013; Cortes, 2010).

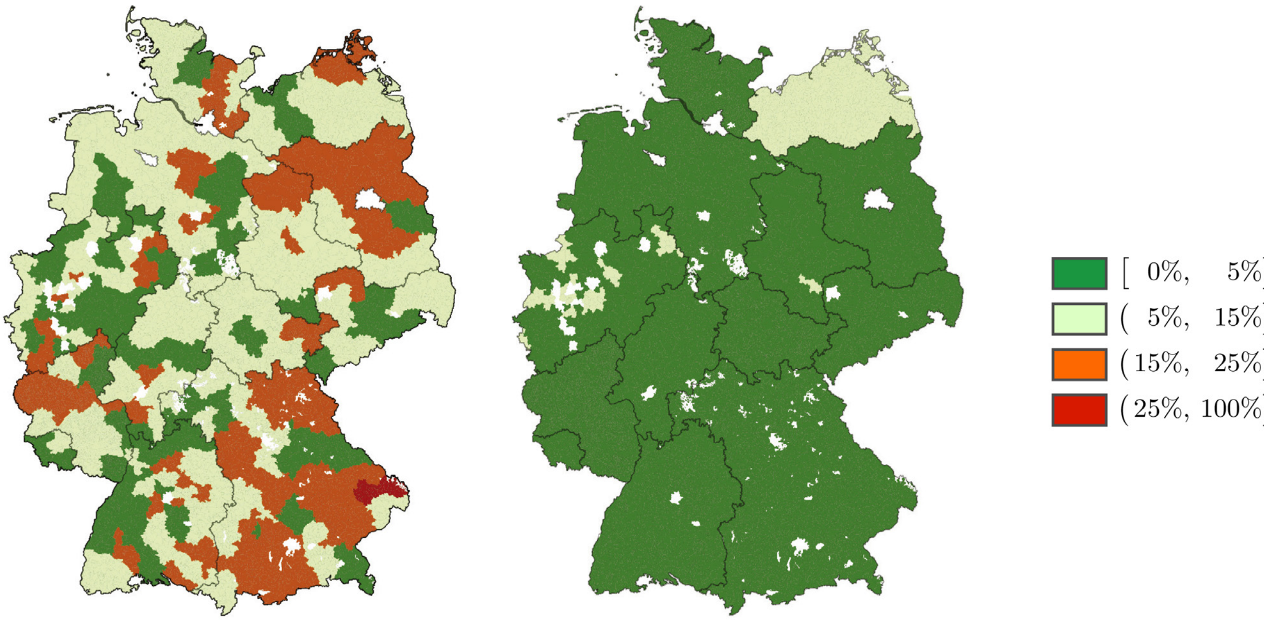
We will present a general unified theory which is based on the relation of constrained geometric clustering and diagrams. In Sections 2 and 3, we analyze the model and prove various favorable properties.

Using several types of diagrams in different spaces, we obtain partitions that are optimized with respect to different criteria: In Euclidean space, we obtain clusters that are particularly well consolidated. Using locally ellipsoidal norms, we can to a certain extent preserve originally existing structures. In a discrete metric space derived from a graph that encodes an intrinsic neighboring relation, we obtain assignments that are guaranteed to be connected. In the theoretical part the various different issues will be visualized with the help of a running example.

Electoral district design. Our prime example will be that of electoral district design which has been approached from various directions over the last half century (see Kalcsics, 2015; Ricca, Scozzari, and Simeone, 2013, and Tasnádi, 2011 for surveys, Goderbauer, 2014 for the example of Germany, and Hwang and Rothblum, 2012; 2013 for general accounts on partitions). Municipalities of a state have to be grouped to form electoral districts. The districts are required to be of nearly equal population and of "reasonable" shape. Hence a crucial nature of the electoral district design problem is that there are several partly conflicting optimization criteria such as the grade of population balance, consolidation, or a desire for some continuity in the development of districts over time. Therefore we will show how our unified approach allows the decision maker to compare several models with different optimization foci.

* Corresponding author.
E-mail addresses: andreas.brieden@unibw.de (A. Brieden), gritzmann@tum.de (P. Gritzmann), klemm@ma.tum.de (F. Klemm).

http://dx.doi.org/10.1016/j.ejor.2017.04.018
0377-2217/



(a) Deviations of the 2013 election districts in Germany (b) Deviations resulting from our methodology. (c) Colors depict the deviation from the federal average district size.

Fig. 1. Deviations from the average population size per district.

Section 4 will show the effect of our method for the federal elections in Germany. The German law (Federal Elections Act, 1993) requires that any deviation of district sizes of more than 15% from the federal average is to be avoided. As a preview, Fig. 1 contrasts the occurring deviations from the 2013 election with the deviations resulting from one of our approaches. The federal average absolute deviation drops significantly from 9.5% for the 2013 election to a value ranging from 2.1% to 2.7% depending on the approach. For most states, these deviations are close to optimal since the average district sizes of the states i.e., the ratios of their numbers of eligible voters and districts differ from the federal average already about as much. See Section 4 for detailed results and the Appendix for further statistics. Furthermore, an online supplement depicts the results of all approaches for the full data set, see http://www-m9.ma.tum.de/material/districting/.

Constrained clustering via generalized Voronoi diagrams. In accordance with (Edelsbrunner & Seidel, 1986), the *generalized Voronoi diagram* for given functions $f_i : \mathcal{X} \to \mathbb{R}$, $i = 1, \ldots, k$, is obtained by assigning each point $x \in \mathcal{X}$ to a subset C_i of $\mathcal{X}$ whose value $f_i(x)$ is minimal. We are interested in clusterings of X that are induced by such diagrams (cf. Section 2.2).

Of course, in order to obtain suitable diagrams, the choice of the functions f_i is crucial. For parameters $(\mathcal{D}, h, \mathcal{S}, \mathcal{M})$ we define the k-tuple of functions $\mathcal{F}(\mathcal{D}, h, \mathcal{S}, \mathcal{M}) := (f_1, \ldots, f_k)$ via

$$f_i(x) := h(d_i(s_i, x)) + \mu_i.$$

Here, $\mathcal{D} := (d_1, \ldots, d_k)$ is a k-tuple of metrics (or more general distance measures) on $\mathcal{X}$, $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is monotonically increasing, $\mathcal{S} := (s_1, \ldots, s_k) \in \mathcal{X}^k$ is a k-tuple of points in $\mathcal{X}$, and $\mathcal{M} := (\mu_1, \ldots, \mu_k) \in \mathbb{R}^k$ is a vector of reals. If the metrics d_i are not all identical, we call the resulting diagram *anisotropic*.

We consider an exemplary selection of types of generalized Voronoi diagrams (see also Ash & Bolker, 1986; Edelsbrunner & Seidel, 1986; Okabe, Boots, Sugihara, & Chiu, 2009; Okabe & Suzuki, 1997). For each of the considered types, Fig. 2 depicts an exemplary diagram together with its induced clustering.

In the Euclidean space, the choice

$$f_i(x) := \|x - s_i\|_2^2 + \mu_i$$

yields *power diagrams*; see (Aurenhammer, Hoffmann, & Aronov, 1998; Barnes, Hoffman, & Rothblum, 1992). For the particular case of *centroidal* diagrams, in which the sites coincide with the resulting centers of gravity of the clusters, the inherent variance is minimized. This can be achieved by optimization over $\mathcal{S}$ (cf. Borgwardt, Brieden, & Gritzmann, 2017; Brieden & Gritzmann, 2010; 2012; Fryer & Holden, 2011).

The setting

$$f_i(x) := \|x - s_i\|_2 + \mu_i$$

yields *additively weighted Voronoi diagrams*.

Allowing for each cluster an individual ellipsoidal norm yields *anisotropic* Voronoi and power diagrams, respectively. Appropriate choices of norms facilitate the integration of further information such as the shape of pre-existing clusters in our application.

We also consider the discrete case $\mathcal{X} = X$. Here, we are given a connected graph $G := (X, E, \delta)$ with a function $\delta : E \to \mathbb{R}_{>0}$ assigning a positive distance to each edge. With $d_G(x, y)$ defined as the length of the shortest x-y-path in G w.r.t. δ, this induces a metric on $\mathcal{X}$. The choice of

$$f_i(x) := d_G(s_i, x) + \mu_i$$

then leads to *shortest-path diagrams*. Such diagrams guarantee the connectivity of all clusters in the underlying graph. This allows to represent intrinsic relations of data points that cannot be easily captured otherwise.

As we will see, the parameters $\mathcal{D}$ and h mainly determine the characteristics of the resulting diagrams. The points s_i then serve as reference points – called *sites* – for the clusters.

It is shown that for any choice of $\mathcal{D}$, h and $\mathcal{S}$ there exists a choice of the additive parameter tuple $\mathcal{M}$, such that the induced clusters are of prescribed weight as well as optimally consolidated (cf. Corollary 2). Thus, we distinguish between the *structural parameters* $\mathcal{D}, h$ and $\mathcal{S}$ and the *feasibility parameter* $\mathcal{M}$. Our approach does not automatically yield integral assignments in general but may require subsequent rounding. However, the number of fractionally assigned points and thus the deviation of cluster weights can be reasonably controlled (see Theorem 5).

Typically, $\mathcal{D}$ and h are defined by the specific application as it determines the requirements on the clusters. One can still optimize

Please cite this article as: A. Brieden et al., Constrained clustering via diagrams: A unified theory and its application to electoral district design, European Journal of Operational Research (2017), http://dx.doi.org/10.1016/j.ejor.2017.04.018