# Enhancing Value by Optimal Segmentation: Application to Determine Newspaper Zones

**Anuj Mehrotra**
Department of Management Science
University of Miami
Coral Gables, FL 33124-8237
e-mail: anuj@miami.edu

**Joseph Shantz**
The Arizona Republic
200 East Van Buren Street
Phoenix, AZ 85004
email: joseph.shantz@pni.com

**Michael A. Trick**
Graduate School of Industrial Administration
Carnegie Mellon University
Pittsburgh, PA 15213-3890
e-mail: trick@cmu.edu

July 31, 2001

**Abstract**

Community sections of newspapers are an increasingly important source of advertising revenue and readership for large city newspapers. Determining the optimal delivery zones for these sections is a complicated optimization problem needing to handle a number of operational issues. We develop an optimization based procedure to generate geographically contiguous zones comprising of customers with similar interests and potential buying behavior. Our methodology is implemented for *The Arizona Republic* and our results demonstrate a 18% to 56% improvement in zones designed over the usual intuitive heuristics. Given the magnitude of the revenue generated by larger local newspapers through zoning, our method demonstrates how optimal zoning can significantly enhance the value of segmentation.

*Key Words: Partitioning, Clustering, Segmentation*

## 1 Introduction

Community sections of large city newspapers are an increasingly important source of revenue and customer loyalty for newspaper publishers. Defining the geographic delivery zones for these sections is a difficult optimization problem with clear implications for the viability of the community section. Good boundaries give a relatively homogeneous readership with similar interests and demographics that lead to a coherent product attractive to advertisers; poor boundaries lead to inconsistent readership and unattractive advertising possibilities. But operational considerations limit feasible zones for the sections. Relatively large areas must receive the same section in order to match the distribution system. Other operational limits involve the number of community sections available and minimum and maximum sizes of service zones.

Our focus is on market segmentation in the face of these operations limitations. There is an extensive literature on market segmentation (see, for instance Smith [19] and Haley [12] for some early work and Arabie and Hubert [1] for a recent survey). Key to this work are creating algorithms for data clustering, an area that has extensive literature in a wide variety of fields (see Jain and Dubes [13] for applications in biology, psychiatry, psychology, archeology, and many other areas). For a recent survey of data clustering algorithms, see Jain, Murty, and Flynn [14]. In that paper, dozens of approaches to creating clusters are presented, with the "k-means" (McQueen [16]) perhaps the best known.

This work differs from previous work in two key ways: first, the application requires contiguous segments based on zip code adjacency. This restriction makes most existing clustering methods inapplicable. Second, our approach stresses finding optimal combinations from a list of possible clusters. Most existing methods are heuristic based. Our approach limits the size of problem we can solve (the approach we propose would not be feasible for thousands of data points), but does find the best combination of clusters to segment the data.

### 1.1 Issues in community sections of local daily newspapers

While the local daily newspaper has been the primary source of information consumption in the household for many years, newspapers have recently seen its readers defect to information providers offering very localized and/or specialized kinds of information. This has become a more serious trend recently, as there is a real consumer preference for localized news.

Most local daily newspapers are able to offer different community sections to meet this consumer demand for community-specific news. These sections generally consist of a separate section, or tab, that is printed and folded in a consumer's newspaper on the press.

The number of sections that can be printed are usually determined by: press capacity to run community sections, the number of consumers that will be served by this information, staff capacity to write community-focused articles, and advertising revenue payback by adding a new community section.

Because these four constraints are the limiting factors of how many community sections can run, newspapers develop and deliver community sections based on pre-determined zones.

### 1.2 Zoning issues

Once the number of community sections are determined through the four constraints listed above, zones are developed based on two new constraints: distribution capacity, and clustering a relatively compact number of consumers in a zone. While outlying zones in a city's newspaper tend to be not as compact as zones that are within the city central, zones are also drawn considering community boundaries. Typically, one or more zip codes are assigned to each zone.

In the past, community zoning was one of the first ways that local advertisers could geographically target an advertising message to consumers. Zoning was not only important to the small, local retailer who want to have their ad placed in areas, but also to multi-store local and national advertisers who sometimes wanted to achieve advertising efficiencies.

Providing zone advertising to smaller advertisers is also a traditional way for newspapers to build relationships with smaller advertisers who may one day grow into multi-store accounts who may eventually need run-of-press advertising (advertising in major sections across the entire market).

Neighborhoods, for a variety of historical and cultural reasons, often attract a differing mix of inhabitants. Targeting consumers based on where they live has been a successful way of marketing products and services to consumers for many years, and many companies provide segmentation tools based on US Census data. For example, a high-end jewelry store would have the most success marketing to wealthier consumers, even if their location is not in the geographic vicinity of wealthier consumers. Newspaper zoning, therefore, offers an additional advantage of targeting consumers who are likely to buy advertisers' products and services that fit the right consumer profile. Therefore, it is essential that newspaper zones provide meaningful clusters of consumers to its advertisers.

Newspaper zone revenue generates millions of dollars for larger local newspapers, and can be as much as 20% of display and classified revenue. In the next section, we develop an optimization based methodology to determine appropriate zones for delivery of community specific newspaper sections. In Section 3, we consider a specific case of *The Arizona Republic* (TAR) newspaper. In Section 4, we present the results and illustrate the benefits of implementing our methodology for TAR. We conclude in Section 5 with some ideas for further research.

# 2 Optimization Based Methodology

## 2.1 An Optimization Model

We model the newspaper zone design problem on a pair of graphs, where each zip code corresponds to a node in each graph, and the the measure of difference between consumers between two zip codes is modeled on the edge between the corresponding nodes in the first complete graph, while the second graph captures the adjacencies of zip codes by including edges between corresponding nodes.

The problem we consider may be stated in graph theoretic terms as follows: Given a pair of graphs $G_1(V, E_1)$ and $G_2(V, E_2)$ where $V$ is the common vertex set in both $G_1$ and $G_2$, $E_1$ and $E_2$ are the edge sets; with edge weights $w_e, e \in E_1$, we consider clustering problems that involve partitioning $V$ into $K$ clusters such that the sum of the weights within the subgraphs induced by $V$ in $G_1$ is minimized (or, equivalently, the total weight on edges between clusters in maximized) and such that the subgraphs induced by $V$ in $G_2$ are connected.

In market segment applications, the weights on the edges $w_e$ represent the similarity between associated zones. There is an extensive literature on how to create these weights (see De Soete [5], Arabie and Hubert [1], and Carmone, Kara, and Maxwell [3] for examples). Our approach applies to any distance measure where the goal is to minimize the sum of the distances within a cluster.

We also incorporate additional restrictions such as the minimum (maximum) number of nodes in each cluster formed. Several graph partitioning type problems have been studied in the literature. Uncapacitated versions of this problem that do not require that the clusters formed be connected model a common problem in qualitative data analysis of partitioning a number of items into clusters, where items within a cluster tend to be homogeneous with respect to some given characteristics. These clusters provide a succinct summary of the data in a way that lends itself to further analysis (for instance, restricting analysis to items in the same cluster, or finding representative items within each cluster). See for example, Chopra and Rao [4], and, Grötschel and Wakabayashi [10, 11]. In the connected versions of the problems we consider, the nodes that can be present in each cluster must induce a connected subgraph. Another version of this problem in which the size of the clusters formed is limited by additional restrictions arises in the context of political districting, see [17].

Statistical methods that are typically used for clustering applications are unable to maintain contiguity of clusters and do not take the graph structure into consideration. Similarly optimization methods [6, 7] developed for solving clustering problems have considered problems that do not have this contiguity requirement. These problems therefore require a new approach. A direct integer programming formulation of this problem is also difficult due to the contiguity restrictions on the clusters. Instead, we consider the following implicit formulation.

## 2.2 An Integer Programming Formulation

Let $s$ be a generic cluster of $V$, and let $x_s$ be a binary variable that will be 1 if cluster $s$ is used in the solution to the clustering problem and 0 otherwise. Let $w_s$ be the weight of the cluster where $w_s$ denotes the sum of weights on edges with both end points in $s$ in the graph $G_1$. Then the clustering formulation that partitions the nodes into a given number, $K$, of zones, is:

**Formulation CF-Partitioning:**

$$\begin{aligned} \textit{Minimize} \quad & \sum_s w_s x_s & & \\ \textit{Subject to} \quad & \sum_{s:i \in s} x_s = 1, & & \textit{for each } i \in V \\ & \sum_s x_s = K & & \\ & x_s \in \{0,1\}. & & \end{aligned}$$

The constraints model that node $i$ must be in exactly one of the chosen clusters and that there must be exactly $K$ clusters chosen. This formulation captures other types of capacitated as well as uncapacitated clustering problems with or without the contiguity restrictions. The contiguity requirement is incorporated by excluding variables corresponding to clusters that are not contiguous.

Despite its appeal, CF-Partitioning is not easy to use for solving the problem because of two issues. First, the number of potential clusters (and therefore the number of variables in the formulation) can be an exponentially large function of the number of zip codes to partition into zones. Second, for any zones that are generated, we must make sure that they are contiguous. The first issue has been addressed in the literature by development of a branch-and-price method for solving problems. Branch-and-price [2] is a method that considers only a small fraction of all possible clusters to implicitly determine an optimal solution. The method uses a technique of *column generation* that iteratively generates feasible clusters that can potentially improve the current solution to the linear programming relaxation of the model. The method of branch-and-price also has to include branching rules that are application dependent and valid in the context of not working with only a partial set of variables. For detailed discussion of such a method for clustering problems, see Mehrotra and Trick [18]. In [18], Mehrotra and Trick demonstrate that a branch-and-price method specifically designed for solving clustering problems outperforms other methods in the literature and represents the state-of-the-art. Even such state-of-the-art methods are unable to solve large clustering problems optimally. In fact, the biggest problems reported solved in the literature have been on graphs with 61 nodes. Additionally, the clustering problems solved in the literature to optimality typically have not needed the contiguity restriction on clusters. Enforcing contiguity requirement on the clusters generated in a branch-and-price can further complicate the methods for finding improving clusters. Given these difficulties and the fact that the problems we wish to solve in this paper are about twice as large as the biggest problems solved to optimality in the literature (and therefore significantly more challenging), we did not pursue a branch-and-price method in this paper. Instead we focused on heuristic methods to generate a pool of potentially attractive zones and solving the integer program restricted to the corresponding set of variables.

## 2.3 Generating Potential Zones

We use two heuristic procedures to generate potential zones.

**Procedure Gen-Node[i, k, U]:** This procedure is used to generate clusters of $k$ nodes that include node $i$ and have a total cost at most $U$. This procedure starts at node $i$ and recursively adds more nodes to the the forming cluster while maintaining contiguity. At any step, each node outside the forming cluster that is adjacent to one of nodes in the forming cluster provides a candidate forming cluster. Any candidate cluster whose total cost exceeds $U$ is abandoned.

We used this procedure to generate clusters containing 2 to $\mid V \mid$ nodes starting at each node of the graph.

**Procedure Gen-Edge[(i,j), k, U]:** This procedure is used to generate clusters of $k$ nodes that include the end nodes of the edge $(i, j)$ and whose total cost is no more than $U$. Again, this recursive procedure starts with one forming cluster containing nodes $i$ and $j$ and adds more nodes to the the forming cluster while maintaining contiguity. At any step, each node outside the forming cluster that is adjacent to one of nodes in the forming cluster provides a potential forming cluster. Any candidate cluster whose total cost exceeds $U$ is abandoned.

We used this procedure to generate clusters containing 2 to $\mid V \mid$ nodes starting at each edge of the graph.

### 2.4 Implementation Issues

**Duplicate Clusters:** The repeated use of the Gen-Node and Gen-Edge procedures produces several duplicate clusters. Before using the clusters as input to the CF-Partitioning model, we removed any duplicate clusters.

**Formulation CF-Covering:** Given the limited number of clusters that we generate using the heuristic methods outlined above, we found that it was not always possible to obtain a feasible solution to the integer program CF-Partitioning. Instead of concentrating on generating more clusters by expanding on the heuristic ideas of building good clusters (which can lead to an exponentially larger number of clusters to generate and consider), we took a more appealing approach. Instead of restricting each zip code to be contained in only one zone, we relaxed the equality constraints to inequality constraints and used the following CF-Covering formulation to determine a set of $K$ zones that contained all the zip codes with a possibility that some zip codes were assigned to multiple zones. Next, we explain how to use the information obtained by a solution to CF-Covering to generate a few additional candidate zones that help solve CF-Partitioning.

$$
\begin{array}{llcll}
\textit{Minimize} & \sum_s w_s x_s & & & \\
\textit{Subject to} & \sum_{s: i \in s} x_s & \geq & 1, & \textit{for all } i \in V \\
& \sum_s x_s & = & K & \\
& x_s & \in & \{0, 1\}. &
\end{array}
$$

**Additional zones:** When a zip code is assigned to multiple zones, it is necessary to generate new zones that might be combined to cover each zip code exactly once. Sometimes it is possible to simply remove the zip code from a zone to create a new zone. This is not always possible because each of the zones that contained that zip code could become non-contiguous on it's removal. Starting with any zone in the solution to CF-Covering that contained any zip codes that were multiply covered, we generated all subsets of that zone that could be obtained by removing one or more multiply covered zip codes. Some of these resulting zones were contiguous and provided potentially improving zones.

Additionally, in order to provide a better chance for CF-Partitioning to have a feasible solution, we used the procedures Gen-Node and Gen-Edge to generate more clusters which did not contain the zip codes covered more than once. This was achieved by simply modifying the adjacency graph by deleting the edges incident to the corresponding nodes.

With these additional clusters, we attempted to solve the formulation CF-Partitioning and found that this was sufficient to obtain a solution to the CF-Partitioning formulation that was almost as good in objective value as the solution to CF-Covering.

# 3 The Arizona Republic Case

*The Arizona Republic* is Arizona's leading provider of news and information, and has published a daily newspaper in Phoenix for more than 110 years. TAR is one of the fastest-growing major metropolitan newspapers in the country. The Republic ranks 15th among U.S. dailies with a daily circulation of 440,000 and a Sunday circulation of 550,000. It's the state's largest newspaper, reaching 1.5 million readers per week.

TAR currently delivers a zoned product in the newspaper within at least some zones 5 times per week. Consumers receive one of 13 different community sections, corresponding to the existing 13 zones, depending on where that consumer receives the paper. The current zones are shown in Figure 1. These zones were configured in 1990, which presents some issues with respect to consumer segmentation.

Arizona has had the second fastest growth rate over the past ten years and the state's population gains trailed only California, Texas, Florida and Georgia. The equivalent of Nebraska moved to the state in the past ten years. Maricopa County represents 59.9% of the state population and grew at a compound rate of 12% - faster than the state. Over the next one to two years, the consensus is that population growth, while healthy, will reflect gains in the 2.8% range compared to the 3.8% compound annual gains achieved in the 1990's, see Table 1.

| | 1990 | 2000 | % chg | 10 CAGR | 2001 | % chg | 2002 | % chg |
|---|---|---|---|---|---|---|---|---|
| Arizona | 3665339 | 5130632 | 40.0 | 3.4 | 5274289 | 2.8 | 5421969 | 2.8 |
| Maricopa County | 2122101 | 3072149 | 44.8 | 3.8 | 3158169 | 2.8 | 3246597 | 2.8 |

Table 1: Population Growth

New-Movers, adults living in their residence less than one year, reflect a total population of 360,000 Valley area adults or 16% of the adult population. They have grown at a compound rate of 4.8% over the past five years exceeding the Valley-wide adult growth rate by 26%. Moreover, recent data suggest that despite the current economic downturn, Arizona continued to experience strong migration-in.

Dynamic growth has also driven dynamic changes in the demographic makeup of Maricopa County. Most of the growth in Maricopa County has been around its periphery with many of its zip codes in states of fast transition from low household income areas to high household income areas, and vice-versa. Because of the changes in the market, zoning needs to be re-examined. This prompted the TAR to do some market research to get a better understanding of the consumer interests.

### 3.1 Market Research Survey

Between February 20 and March 24, 2001, a market research survey was conducted based on 1,311 telephone interviews with adult residents in the Phoenix MSA, averaging 24 minutes in length. A randomly selected sample of residential phone numbers was generated to include new telephone listing and adults with unlisted phone numbers. Interviewing was conducted by Schulman, Ronca, and Bucuvalas of New York. With respect to how this survey pertains to this project, respondents' interest levels across 56 different types of newspaper content were measured.

The interests that were measured are contained in Appendix A. Interests were measured on a 3-point scale - from very interested to somewhat interested to not interested.

One of the issues that came out in this study is that 3 of the top 4 topics that were listed by both readers and non-readers dealt with issues that were local within their community in the Phoenix MSA. The top 3 interests of non-readers of the Arizona Republic are "Good things happening in your area," (85% top 2 box) "Places to go and things to do," (85% top 2 box) and "News from your own town or city" (82% top 2 box.) For regular readers of TAR, these percentages were higher: 95%, 92%, and 92%, respectively.

While consumers report that they want information about things that are local to them, consumers also have different information needs at different times in their life. For example, a household with children present is going to be more interested in information about schools and where to take the children for entertainment on the weekends. Similarly, income is a driver of different kinds of information needs. Households with a very high income are more likely to be interested in financial issues and high end leisure activities; while households with low income tend to have interests in different kinds of leisure activities. Finally, the ages of the people in the household also have a great deal to do with the kinds of information that these consumers will be interested in. Young adults aged 18-24 are interested in where to start their life - they need information on jobs and where to live, and where they can go for entertainment. Senior citizens, on the other hand, are ! more interested in health issues, etc.

For this project, we chose these three axes that determined where consumers were clustered: age, income, and presence of children.

### 3.2 Data for the Optimization Model

The opportunity that TAR management is exploring is to rezone the community sections by clustering consumers based on the 56 readership interest measurements. Because zones contain a unique set of adjacent zip codes that cover a market, there needed to be a way to transform the market research data from a respondent matrix of interests into a zip code matrix of interests. Because many zip codes only had 3 or 4 respondents, extrapolating readership interest based on this low number of respondents is statistically erroneous.

However, TAR has an extensive consumer database of the entire state of Arizona. This enabled matching respondent telephone numbers to the consumer database. From this point, lifestyle clustering information could be attached back to the respondents in the research (see Appendix B for detail on segment types). From this point, respondents were assigned to segments based on household income, age of the respondent, and presence of children. Respondents were tagged with a field indicating to which segment the respondent belonged.

A matrix of interests for a segment group was generated by taking the average score across all respondents across all interest measurements. Next, an interest matrix at the zip code level was generated by taking the weighted average of the segment interest matrices based on the total population of a particular segment in that zip code. This matrix was then used for generating the weights on the edges of the graph on which the problem would be modeled. The weight on an edge quantifies the difference between the zip codes that correspond to the end nodes as explained in Section 2. This was enabled by taking the sum of the absolute value of the difference between one zip code and another across each of the interest measurements to generate a 118 x 118 matrix, corresponding to the absolute value of the difference between one zip code and another across all 118 zip codes in the market area. Additionally, another 118 x 118 adjacency matrix was developed to indicate the adjacencies between the 118 zip codes.

# 4 Computational Results

TAR presently uses 13 delivery zones as displayed in Figure 1. The sum of the edge weights in each of the 13 zones, and the number of zip codes in each zone are summarized in Table 2.

The number of zip codes vary between 3 and 14 per zone and the weights of the zones vary between 2.4 and 129.5 with the total weights across all zones summing up to 648.9.

We implemented our Gen-Node and Gen-Edge procedures as outlined earlier to generate potential zones. TAR management also specified that any zone must at least contain 2 zip codes. Hence, we restricted our attention to zones with at least 2 zip codes. We used 500 as an upper bound on the weight of any one zone in our procedure. This resulted in generation of 4374 zones. An optimal solution to CF-Covering with $K = 13$ using these candidate zones resulted in six zip codes that were in multiple resulting zones. We generated additional zones that did not include these six zip codes by using the procedures Gen-Node and Gen-Edge on adjacency graphs modified to exclude edges adjacent to the nodes corresponding to these 6 zip codes. After omitting any duplicate zones, this yielded 5816 potential zones. Even with these 5816 zones, a new optimal solution to CF-Covering with $K = 13$ had two zip codes belonging to two zones. We generated all contiguous subsets of the zones that excluded one or more of these zip codes. This yielded another 16 potential zones for a total of 5832 candidate zones. For all the results that are described next, we solved the CF-Partitioning formulation using these 5832 candidate zones. We used CPLEX (version 6) on a DEC ALPHA 3000 (Model 900) for solving the CF-Partitioning formulations.

## 4.1 Solution Quality and Number of Zones

While TAR currently employs 13 zones, there is interest in determining configurations with more zones to understand the additional benefits of finer segmentation of customers. However, operational constraints and increase in costs make printing over 18 different community sections unrealistic. Hence, we explored finding solutions which partitioned the 118 zip codes into 13 to 18 zones. To quantify the marginal benefits possible by using an additional zone, we solved the CF-Partitioning Model for $K = 13, ..., 18$. As might be expected, the results show that much better results are obtained as the number of zones used is increased. Table 3 shows the objective value obtained and the cputime spent for 13 through 18 zone configurations. We also list the percentage savings based on the current 13 zone configuration. The 13 zone proposal shows an 18% improvement and the 18 zone proposal shows as high as 55% improvement. *The Arizona Republic* is presently considering the feasibility of implementing the 18 zone proposal and analyzing cost tradeoffs and the implications on potential revenue that additional zones may be able to generate.

## 4.2 Proposed 13 zone solution

Figure 2 depicts the 13 zone proposal that was generated through the use of our methodology. The characteristics of the individual zones in our 13 zone solution are summarized in Table 4. The

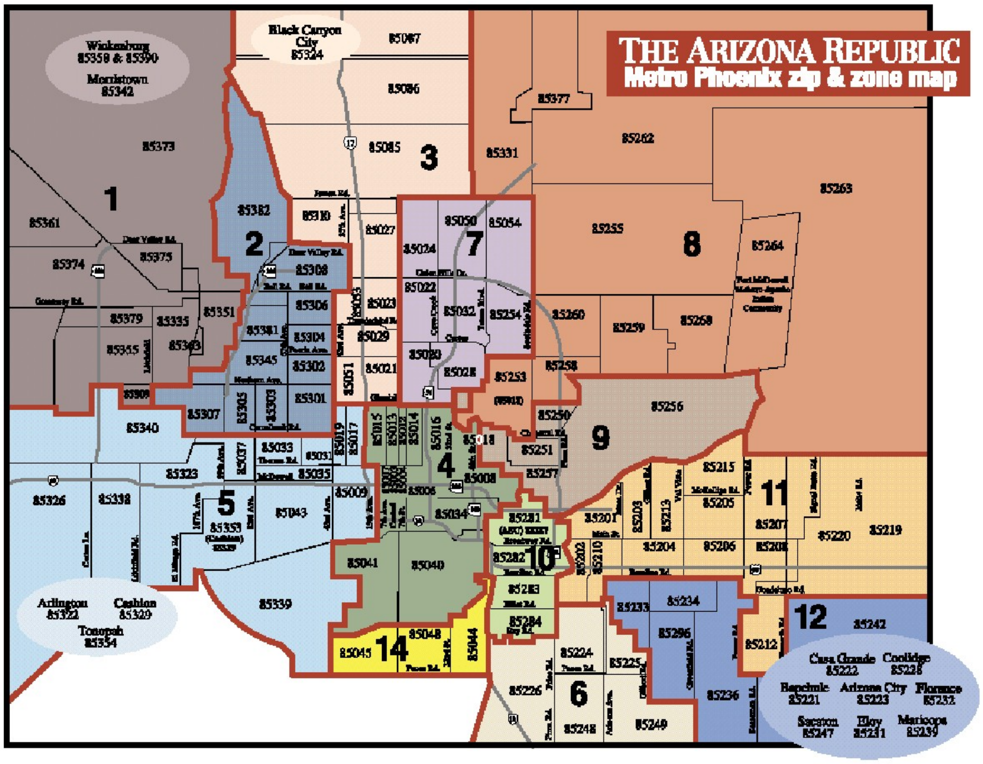


Figure 1: Currently Employed Zones by TAR

| Zone | Weight | Zip Codes |
|---|---|---|
| 1 | 129.5 | 12 |
| 2 | 46.2 | 11 |
| 3 | 94.3 | 12 |
| 4 | 54.2 | 13 |
| 5 | 75.6 | 14 |
| 6 | 8.2 | 5 |
| 7 | 24.9 | 8 |
| 8 | 61.2 | 11 |
| 9 | 9.1 | 5 |
| 10 | 6.2 | 4 |
| 11 | 112.8 | 14 |
| 12 | 24.3 | 6 |
| 13 | 2.4 | 3 |
| Total | 648.9 | 118 |

Table 2: TAR's currently employed 13 zones

| | Number of zones | | | | | |
|---|---|---|---|---|---|---|
| | 13 | 14 | 15 | 16 | 17 | 18 |
| cpu seconds | 28.69 | 57.86 | 83.08 | 66.74 | 49.19 | 44.49 |
| Objective Value | 529.13 | 456.08 | 398.47 | 358.78 | 313.03 | 285.96 |
| Percentage Savings | 18.45 | 29.71 | 38.59 | 44.70 | 51.76 | 55.93 |

Table 3: Solutions obtained with our methodology

number of zip codes in these range from 4 to 20 and the weight of zones range from 0.0 to 127.4. These ranges are not too different from the corresponding ranges in the current implementation.

The solution obtained appealed to the TAR management for several reasons and helped understand their consumer profile better. The advantage of this solution is that it compacts readership interest demographically. For example, Southern Phoenix tends to be lower income with a high Hispanic population. Readers in this zone are very interested in news from Mexico, and employment information.

The next large area of note is the zone that consists of north Phoenix east to the area in South Scottsdale. This zone tends to consist of households with Seniors and a high Hispanic family concentration. This zone is more near the midpoint of household income, and therefore issues are slightly different. There is more concern about schools, and leisure activities.

The zones to the immediate east and north of Phoenix including Avondale, Glendale are zones that would not be traditional newspaper community zones, as they split up and cross over multiple municipal lines. However, they are of significant value as the Arizona Republic is able to deliver both content and advertising messages directly to addresses that fit a specific demographic profile. Therefore, this zoning configuration allows advertisers to deliver their message to consumers that would be the best prospects.

Moreover, this zoning configuration is helpful because it demonstrates what current areas in the Phoenix MSA are in transition. This is useful information in determining where new communities are likely to emerge in the future.

### 4.3 Proposed 18 zone solution

Figure 3 graphically depicts the 18 zone proposal that was generated through the use of our methodology. The characteristics of the individual zones in our 18 zone solution are summarized in Table 5. The number of zip codes in these range from 3 to 11 and the weight of zones range from 0.0 to 52.6. These ranges indicate that much better segmentation is possible with this increased number of zones.

The 18 zone proposal shows even more clearly where geographic pockets of interest exist. Areas around central Phoenix that are broken up represent in clearer detail those areas of transition. Because there is demographic diversity in some of these zip codes that can range from poor families to wealthy singles, these micro-zones around Phoenix best describe the eclectic reader interests of the eclectic demographics. These zip codes could easily convert into a predominant cluster of readership interests that would come with a shift to a predominant demographic group. This is not only useful information going forward, but it also is useful to advertisers who have products and services that can serve this unique audiences.

Moreover, newspaper content to these pockets of interests is important to deliver to maintain and grow the reader base of *The Arizona Republic*. Delivering these pockets of interests to advertisers in this way may also be an effective way to grow revenue.

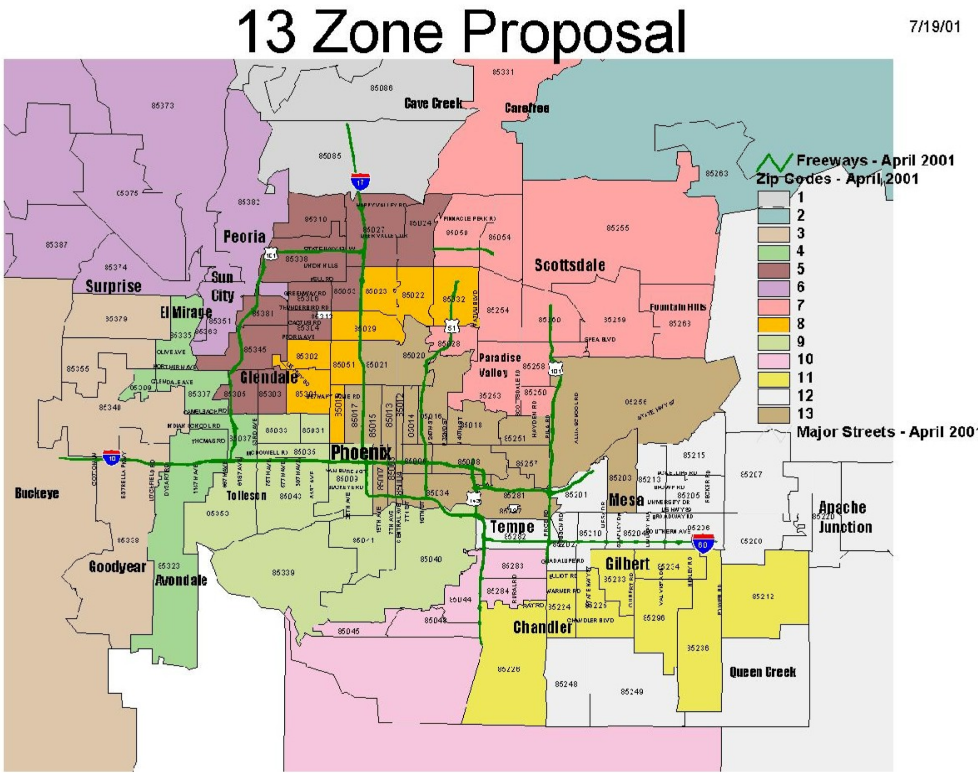


Figure 2: Proposed 13 Zone Solution

| Zone | Weight | Zip Codes |
|---|---|---|
| 1 | 0.0 | 4 |
| 2 | 3.0 | 3 |
| 3 | 9.9 | 5 |
| 4 | 12.7 | 5 |
| 5 | 32.0 | 11 |
| 6 | 52.6 | 9 |
| 7 | 54.4 | 12 |
| 8 | 11.3 | 8 |
| 9 | 20.5 | 9 |
| 10 | 26.8 | 6 |
| 11 | 56.5 | 9 |
| 12 | 121.8 | 17 |
| 13 | 127.4 | 20 |
| Total | 529.1 | 118 |

Table 4: 13 zone solution using our methodology

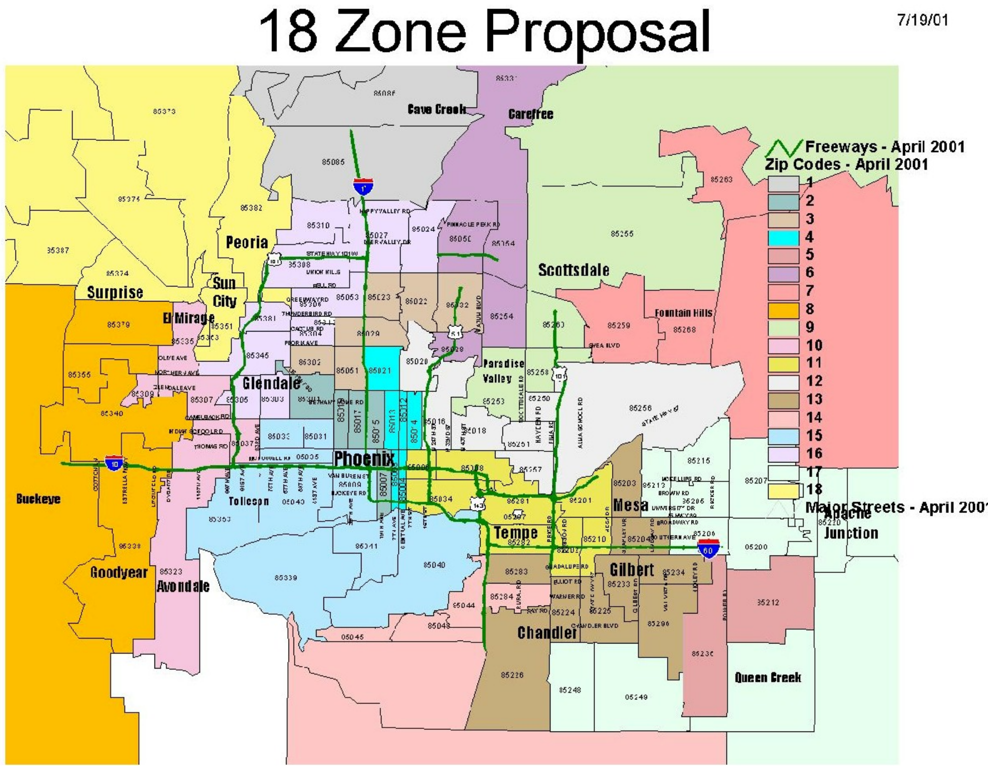


Figure 3: Proposed 18 Zone Solution

| Zone | Weight | Zip Codes |
|---|---|---|
| 1 | 0.0 | 4 |
| 2 | 4.3 | 5 |
| 3 | 4.7 | 6 |
| 4 | 5.3 | 6 |
| 5 | 6.1 | 3 |
| 6 | 6.8 | 5 |
| 7 | 9.7 | 4 |
| 8 | 9.9 | 5 |
| 9 | 11.2 | 6 |
| 10 | 12.7 | 5 |
| 11 | 15.2 | 8 |
| 12 | 15.6 | 7 |
| 13 | 18.7 | 9 |
| 14 | 19.4 | 5 |
| 15 | 20.5 | 9 |
| 16 | 32.0 | 11 |
| 17 | 41.2 | 11 |
| 18 | 52.6 | 9 |
| Total | 285.9 | 118 |

Table 5: 18 zone solution using our methodology

One concern that may need to be addressed is that this zoning configuration splits up Phoenix and Glendale cities into many zones. Additional market research would need to be conducted to learn if readers identify themselves with the municipality they belong to. If this is the case, then addressing the issue of actual content may be problematic within this configuration.

That is, if "news about my community" is an important issue for those in both Glendale zones, then serving the Phoenix city reader that falls within the East Glendale zone may be an issue. From an advertising perspective, the 18-zone proposal is actionable by tailoring advertising messages and the distribution of those messages to these clusters of interest.

## 5 Conclusions and Further Research

We have developed an optimization based procedure to generate geographically contiguous zones comprising of customers with similar interests and potential buying behavior. This work extends the optimization methods for solving clustering problems as the clustering techniques in the literature are unable to deal well with the contiguity requirements on the zones. We have used our methodology to develop delivery zones for community sections of a newspaper for *The Arizona Republic*. The results obtained indicate substantial improvements over the ad-hoc and/or usually employed heuristic procedures. These improvements add significant value by increasing the possibility of targeting consumers who are likely to buy advertisers' products and services that fit the right consumer profile. Given the magnitude of the revenue generated by larger local newspapers through zoning, our method demonstrates how optimal zoning can significantly enhance the value of segmentation. While our methodology has been sufficient to fulfill our goals for this research, several interesting research avenues can be further explored.

For example, our procedures did not attempt to solve the formulations optimally because they only considered a small, albeit attractive subset of all candidate solutions. A natural extension of this work is to develop additional ways to generate attractive zones. For instance, we have experimented with variations of tabu search (see Glover [8, 9]), where we modify our greedy heuristics to "skip" over some close zip codes in order to generate different zones. More complicated heuristics could generate still other zones. Initial results suggest that our greedy heuristics do very well, but can be improved on. Additionally, we did not look for zones to be compact: it turned out that the zones that we generated were geographically compact, but it might be worthwhile to enforce a compactness requirement to avoid long zones that are contiguous but operationally unpalatable.

To determine how well we can do, a branch-and-price technique can be developed for solving these problems to optimality. Whether such techniques will be able to solve the problems of this size remains to be seen (results from Mehrotra and Trick [18] do not reach the size we are interested in here).

Our methodology also opens up possibilities to provide improved segmentation for applications other than designing newspaper delivery zones. Such applications where segmentation benefits product placement, customer satisfaction, identification of bands of interests and improved advertising, for example, can achieve additional value from using this approach. Depending on the specific application, a measure of closeness of interests may need to be modified. Similarly, based on the input from TAR management, we modeled the zip codes as the nodes in our graphs. For other applications, it may be necessary to use smaller geographic units as the starting point, potentially making the resulting problems quite large. In that case, it may be better to focus more on further improving the methods for generating attractive clusters, rather than developing a method for seeking an optimal solution.

## References


[1] Arabie, P. and L. Hubert, 1994. Cluster Analysis in Marketing Research. in *Advanced Methods in Marketing Research*, R.P. Bagozzi, ed. Oxford: Blackwell, 160-189.

[2] Barnhart C., E. L. Johnson, G. L. Nemhauser, M. W. P. Savelsbergh, and P. H. Vance, 1998. Branch-and-Price: Column Generation for Huge Integer Programs. *Operations Research* **46:3**, 316-329.

[3] Carmone,Jr., F.J., A. Kara, and S. Maxwell, 1999. HINoV: A New Model to Improve Market Segment Definition by Identifying Noisy Variables. *Journal of Marketing Research*, **36** 501-509.

[4] Chopra S. and M. R. Rao, 1993. The partition problem. *Mathematical Programming* **59**, 87-116.

[5] De Soete, G., 1988. OVWTRE: A Program for Optimal Variable Weighting for Ultrametric and Additive Tree Fitting. *Journal of Classification* **5**, 101–104.

[6] Ferreira C. E., A. Martin, C. C. De Souza, R. Weismantel and L. A. Wolsey, 1996. Formulations and valid inequalities for the node capacitated graph partitioning problem. *Mathematical Programming,* **74**, 247-266.

[7] Ferreira C. E., A. Martin, C. C. De Souza, R. Weismantel and L. A. Wolsey, 1998. The node capacitated graph partitioning problem: A computational study. *Mathematical Programming,* Series B **81**, 229-256.

[8] Glover, F., 1989. Tabu Search - Part I. ORSA Journal of Computing **1**, 190-206.

[9] Glover, F., 1990. Tabu Search - Part II. ORSA Journal of Computing **2**, 4-32.

[10] M. Grötschel and Y. Wakabayashi, 1989. A cutting plane algorithm for a clustering problem. *Mathematical Programming* **45**, 59-96.

[11] M. Grötschel and Y. Wakabayashi, 1990. Facets of the clique partitioning polytope. *Mathematical Programming* **47**, 367-387.

[12] Haley R., 1968. Benefit segmentation: A decision-oriented research tool. *Journal of Marketing* **32**, 30-35.

[13] Jain, A.K. and R.C. Dubes, 1988. Algorithms for Clustering Data, Prentice Hall, New Jersey.

[14] Jain, A.K., M.N. Murty, and P.J. Flynn, 1999. Data Clustering: A review. *ACM Computing Surveys* **31**, 264-323.

[15] Johnson E. L., A. Mehrotra, and G. L. Nemhauser, 1993. Min-cut clustering. *Mathematical Programming*, **62**, 133-151.

[16] McQueen, J. 1967. Some methods for classificaiton and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 281-297.

[17] Mehrotra A., E. L. Johnson, and G. L. Nemhauser, 1998. An optimization based heuristic for poiltical districting. *Management Science*, **44:8**, 1100-1114.

[18] Mehrotra A. and M. A. Trick, 1998. Cliques and Clustering: A Combinatorial Approach. *Operations Research Letters*, **22:1**, pp 1-12.

[19] Smith, W. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing* **21**, 3-8.

[20] Vance P. H., C. Barnhart, E. L. Johnson, and G. L. Nemhauser, 1993. Solving Binary Cutting Stock Problems by Column Generation and Branch-and-Bound. *Computational Optimization and Applications*, III, 111-130.

## A Interests Measured

Respondents were asked a series of questions on interests. The question began: "How interested are you in reading about ..." The following list contains all of the interest levels across the topics measured

```
A. News from your own town or city
B. Information about local organizations, clubs, and meetings
C. Good things happening in your area
D. Education and the local schools
E. Local police, crime, and court news
F. Growth and development in the local area
G. World and national news
H. National business and financial news
I. News about large businesses in this area
J. News about small businesses in this area
K. Personal finance and investment advice
L. News about Arizona colleges and universities
M. News from around the state of Arizona
N. Arizona state government and legislature news
O. Local and county government news
P. Listings of stocks, bonds, and mutual funds
Q. Editorials about your own town or community
R. Letters to the editor about your own town or community
S. The local environment
T. News from Mexico
U. News from Western Phoenix
V. News from Eastern Phoenix
W. News from Sun City and Sun City West
X. News from Glendale and Peoria
Y. News from Scottsdale and Paradise Valley
Z. News from Chandler
AA. News from Tempe
BB. News from Mesa
CC. The arts including theater, dance, concerts and art exhibits
DD. Food, food trends and recipes
EE. Home repair, home improvement and redecorating
FF. Family life and raising children
```

GG. Health, medicine and fitness
HH. Jobs, career and workplace issues
II. History of the local area
JJ. Computers and the Internet
KK. Science and technology
LL. Places to go and things to do
MM. Restaurants and dining
NN. Movies and movie reviews
OO. Books and book reviews
PP. Outdoor recreation such as hiking, camping and biking
QQ. Local high school sports
RR. The Arizona Diamondbacks
SS. Other Major league baseball
TT. The Phoenix Suns
UU. Other NBA Pro basketball
VV. The Arizona Cardinals
WW. Other NFL Pro Football
XX. The Phoenix Coyotes
YY. Other NHL Pro Hockey
ZZ. Arizona State University sports
AAA. University of Arizona sports
BBB. National college sports
CCC. Auto racing
DDD. Golf

## B Segment clusters

Consumers were clustered by the following data:

```
Age:          18-34, 35-54, 55+
Household Income:  Less than $25,000, $25,000 - $50,000, Greater than $50,000
Presence of Children: present, not present
This provided 18 different consumer clusters.
```