Optimal Realignment of Athletic Conferences

Anthony Erickson | Senior
Business Administration

Abstract

This article presents a mathematical analysis of the current realignment plan for men's, Division I, college hockey. Comparisons are made between existing alignments, proposed alignments, random alignments, and optimal alignments, with respect to various measures. It is shown how proposed alignments do not minimize travel distance nor maximize attendance. It is shown also how a non deterministic clustering procedure can be expected to outperform the proposed alignments, with respect to these measures. Although this clustering procedure is not almost surely optimal, it is shown in our hockey setting to be an effective approximation, being nearly optimal and easily computable. R programs are provided in the appendix.
Keywords: athletic conferences, optimization, k-means clustering

Introduction

The Western Collegiate Hockey Association (WCHA) and the Central Collegiate Hockey Association (CCHA) are two established, men's Division I college hockey conferences, whose teams have claimed a combined 48 national championships [4]. The conferences are facing significant challenges, extinction in the case of the CCHA and significant alteration in the case of the WCHA. This is due to the emergence of the Big Ten ice hockey conference [1] and resulting realignment [5].

There are those who vehemently disagree with the premise that the realignment plans benefit college hockey generally (see [10],[5],[11] or [6]). To give a sense of the associated controversy, a quote from a storied veteran of college hockey seems appropriate. Former player and current coach Dean Blais has said, "We (the college hockey community) didn't decide on this....I don't think it was for the good of hockey [10]." He insinuates, perhaps, that the Big Ten Network is behind the sudden changes to the college hockey landscape.

While some commentators have pointed out certain benefits of the proposed realignment (see [9] or [12]), it seems premature to form an opinion on the matter based mostly on the rhetoric of columnists. Here, in order to contribute to the debate, we present a mathematical analysis of how best to group teams into conferences. Our logical approach assumes the desirability of alignments that minimize total travel distance and maximize attendance. We focus mainly on teams located in the Midwest.

The Teams

As of 2013 there are 59 teams in men's, Division I, college hockey [13]. Since most of the conference realignment involves teams from the WCHA and CCHA, we focus on the Midwest region. We exclude the Alaska teams from our analysis. Since Notre Dame has joined Hockey East [13], we exclude them as well. 21 teams remain, as mapped in Figure 1. The latitude and longitude coordinates were obtained through a Google search, and plotted using "Map-It" [16].

We speak of the traditional alignment into WCHA and CCHA teams as the WCHA-CCHA alignment. We speak of the proposed realignment as the Big Ten alignment. See Table 2.1 for details.

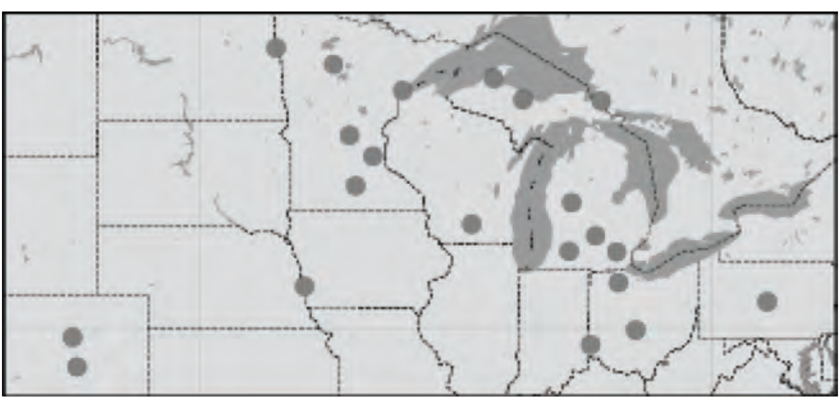

Figure 1: Locations of the universities under consideration

Table 2.1: Proposed realignment of teams [13]. NCHC stands for the newly created National Collegiate Hockey Conference.

University	2012 Conference	2013 Conference
Bowling Green State University	CCHA	WCHA
Ferris State University	CCHA	WCHA
Lake Superior State University	CCHA	WCHA
Miami University	CCHA	NCHC
University of Michigan	CCHA	Big Ten
Michigan State University	CCHA	Big Ten
Northern Michigan University	CCHA	WCHA
Ohio State University	CCHA	Big Ten
Western Michigan University	CCHA	NCHC
Bemidji State University	WCHA	WCHA
Colorado College	WCHA	NCHC
University of Denver	WCHA	NCHC
Michigan Technological University	WCHA	WCHA
University of Minnesota	WCHA	Big Ten
University of Minnesota–Duluth	WCHA	NCHC
Minnesota State University	WCHA	WCHA
University of Nebraska–Omaha	WCHA	NCHC
University of North Dakota	WCHA	NCHC
Saint Cloud State University	WCHA	NCHC
University of Wisconsin	WCHA	Big Ten
Pennsylvania State University	N.A.	Big Ten

Minimizing Travel Distance

In this section we select the alignment that minimizes travel distance between teams. Road distances between universities have been determined using Google Maps.

Definition 3.1. The **distance matrix** is defined to be the 21 by 21 matrix D where D[*i*, *j*] is the distance between the hometown of team *i* and the hometown of team *j*.

Before using these data we should define precisely what is meant by an alignment.

Definition 3.2. An **alignment** is a vector *x*, of length 21, whose entries take a finite set of values that indicate conference membership.

History guides us to only consider alignments into two or three conferences; that is vectors whose entries take only two or three values. For each conference, indexed by *l*, of an alignment into *k* conferences, there is an associated submatrix of *D*, that we denote with D_l. It is an n_l by n_l matrix with entries indexed by *i* and *j*. With such notation the following definition is possible.

Definition 3.3. For an alignment of teams into k conferences, each indexed by l and containing nl teams, define the **travel distance** to be

$$d(\mathbf{x}) = \sum_{l=1}^{k} \frac{\sum_{i=1}^{n_l} \sum_{j=1}^{n_l} D_l[i,j]}{2(n_l - 1)}.$$

The $2(n_l - 1)$ terms are included so that the travel distance gives an approximate measure of the average, weekly, distance traveled by all teams in all conferences over the course of the entire season.

We would like to minimize *d* over all possible alignments. This was initially accomplished by computing *d* for every possible alignment. (For a more elegant but less certain approach see Section 5 on clustering.) We programmed R (see [14]) to count in both base two and base three, and after adding zeros for place holders and recognizing that each number represents an alignment, we then computed the travel distance d for all possible alignments. This took approximately six days of computing time. The associated R program is included in the appendix.

The results indicate that the Big Ten alignment was definitely not chosen to minimize travel distance. Indeed, the travel distance for the Big Ten alignment is not much better than what we would expect from a random alignment. For a comparison of the travel distances for random alignments, the existing alignment, the Big Ten alignment, and the optimal alignment see Figure 2. A map of the optimal alignment that minimizes travel distance is displayed in the fourth plot of Figure 4.

Maximizing Attendance

In this section we select the alignment that maximizes attendance. The starting point toward a rigorous measure of attendance across alignments is a matrix of attendance data.

Definition 4.1. The **attendance matrix** is defined to be the 21 by 21 matrix A where $A[i, j]$ is an estimate for the average historical attendance when team i hosts team j.

A limitation of our analysis is that we were only able to obtain attendance data for the 2011–2012 season. Data were obtained from box scores stored online [15]. If a row team hosted a column team multiple times then an average attendance was entered. If a row team did not host a column team during the 2011–2012 season we left the corresponding entry blank.

Definition 4.2. With M_i denoting the average of the entries in row i of A, define the standardized attendance matrix, $\bar{A}$, to be the 21 by 21 matrix with entries determined by

$$\bar{A}[i, j] = \frac{A[i, j] - M_i}{M_i}.$$

In case of lacking data we set $\bar{A}[i, j] = 0$. We can now define a measure of attendance for alignments. In the following definition the $2(n_l - 1)$ factors are used mainly to allow for adequate comparison of alignments that utilize conferences of differing sizes.

Definition 4.3. For an alignment, x, of teams into k conferences, each indexed by l and containing n_l teams, define the **attendance score** to be

$$a(\mathbf{x}) = \sum_{l=1}^{k} \frac{\sum_{i=1}^{n_l} \sum_{j=1}^{n_l} \bar{A}_l[i, j]}{2(n_l - 1)}.$$

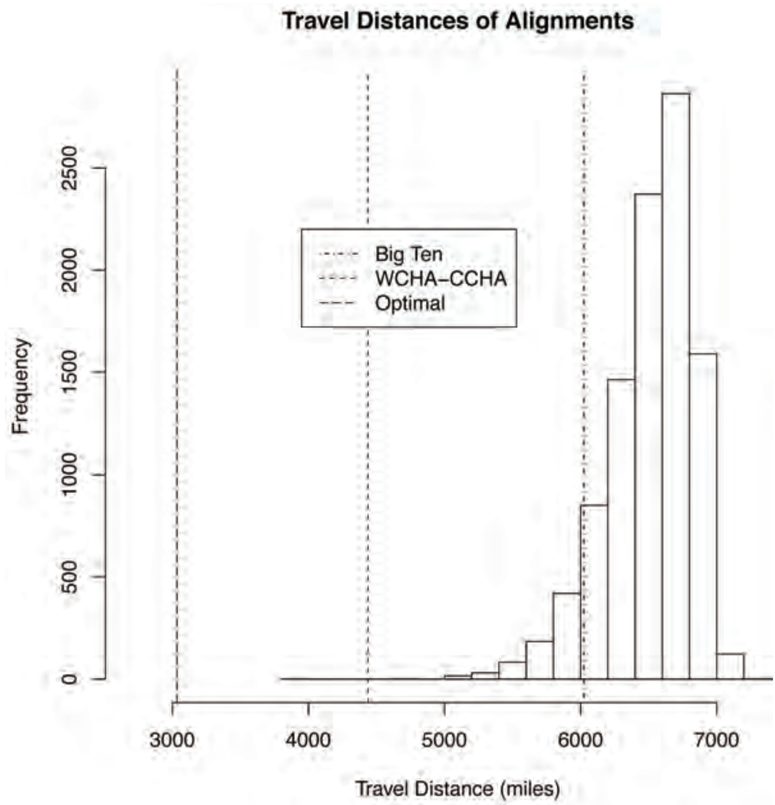

Figure 2: Distances were computed for each of ten thousand, randomly selected alignments, and then plotted in a histogram so as to provide a reference background upon which to compare the distances associated with three different alignments of interest.

As in Section 3, we seek to find an optimal alignment, where this time we are searching for an alignment that maximizes the attendance score *a*. Analogous techniques are used (see the appendix for the R commands). The results are displayed in Figure 3.

Clustering Teams

Here we proceed without pointed objectives such as minimizing travel distance or maximizing attendance. Our aim is to select promising alignments by clustering teams into groups based on a heuristic method. The method is known as k-means clustering, and the mathematical details can be accessed within

Johnson and Wichern's Applied Multivariate Statistical Analysis [8]. All that is required here, however, is a basic understanding of the method. For simplicity we set k equal to 3 so our objective is to suitably cluster the 21 teams (as viewed on a map) into 3 groups.

We start by assigning coordinates to each of the teams hometowns, and for this purpose we employ latitude as the y coordinate and a transformed version of longitude as the *x* coordinate. We store this information in a 21 by 2 matrix, with the x coordinates in the first column and the y coordinates in the second column. We denote the matrix with *C* and its entries with C*[i, j]*.

While it might be argued that within this section the distances between the hometown's of pairs of teams should be measured along geodesics (great circles of the Earth), we instead simply utilize a Euclidean approximation, since the patch of the Earth under consideration is not too large. Remember, we have excluded the Alaskan teams from our analysis. Also, precise attention to detail is not overwhelmingly important here, as our clustering method is heuristic and not even deterministic, as we shall see.

Once x and y coordinates have been assigned to each of the teams, we randomly select three pairs of coordinates to serve as three initial centers for conferences. Teams are then assigned to the conference (center) that is closest—as measured with Euclidean distance in the plane—to their hometown's coordinates. After such assignment is complete, updated centers are computed—for each conference a new center is established as the geometric mean of the coordinates of its (previous) teams. All the teams are then reassigned, resulting in a new alignment. This process is then repeated until repeated iteration no longer changes the selected alignment.

We found through experimentation that ten iterations is typically sufficient to obtain results. We ran thirty iterations just to make sure, and we did this for each of seventy triplets of initial centers, each randomly chosen. The results were eleven attractive clusters of teams into alignments, each of which revealed unforeseen possibilities for possible conference realignment. Some associated plots are on display in Figure 4.

For each of the eleven results of our clustering procedure it is possible to compute an associated travel distance and attendance score. Such vectors of data, and their standardizations can be seen in Table 6.1. The ideal alignment possesses both a low travel distance and a high attendance score. However, in order to select the "best" alignment, some subjectivity is required.

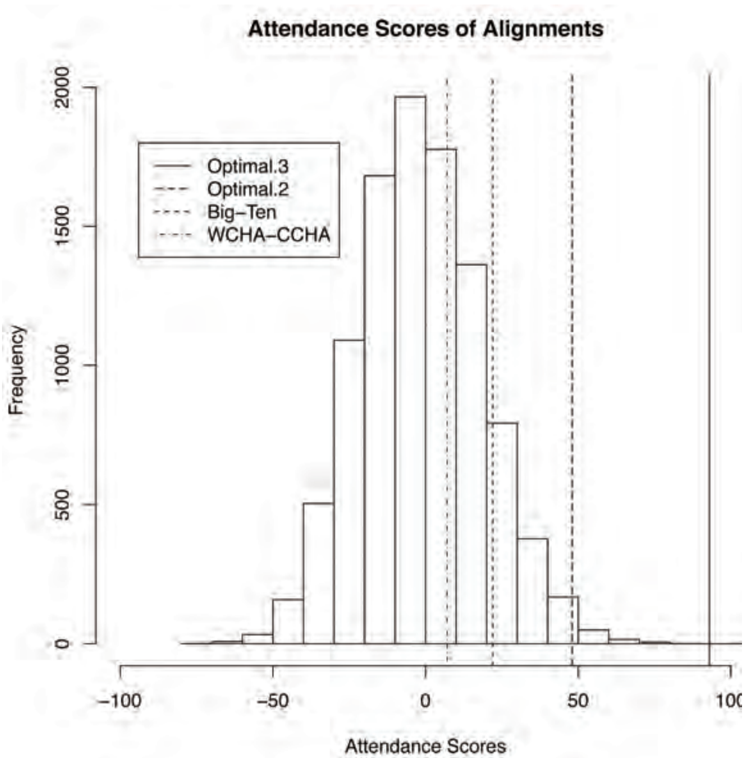

Figure 3: Attendance scores were computed for each of ten thousand, randomly selected, three-conference alignments, and then plotted so as to provide a reference background upon which to compare the the attendance scores associated with various alignments of interest. For k = 2, 3, "Optimal.k" refers to optimal alignment into k conferences.

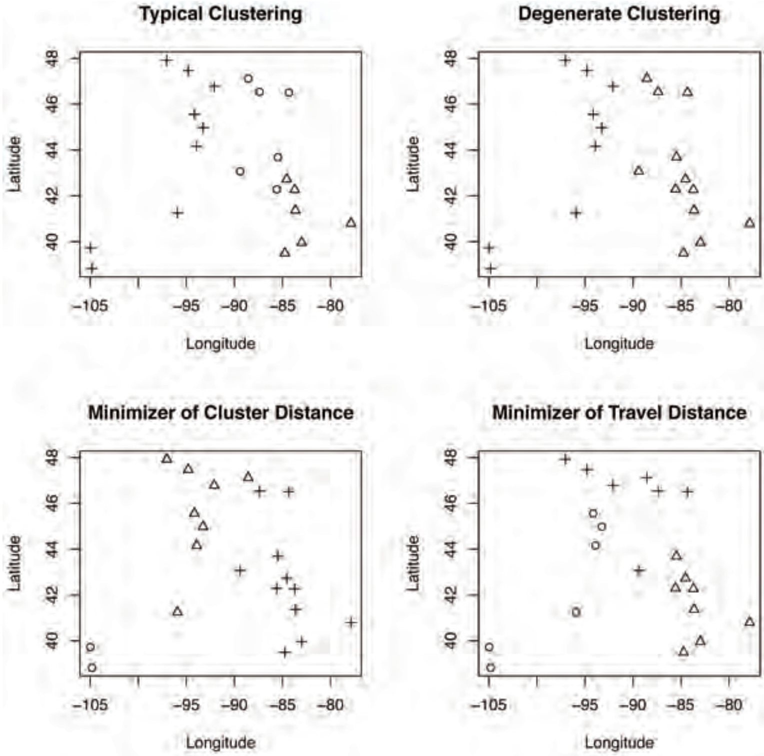

Figure 4: The non deterministic nature of the clustering algorithm is displayed in the top two graphics—different, randomly selected triplets of initial centers give rise to different clusters. After running the algorithm for 70 different randomly selected triplets of initial centers it was possible to select from the 11 different observed outcomes an empirical minimizer of cluster distance. This minimizer is on display above. It is important to note that none of the seventy trials was successful in selecting our previously determined minimizer of travel distance, that is also on display above for comparison.

Conclusions

Based on our investigations, it is clear that the Big Ten alignment was not proposed in order to minimize distance traveled. The evidence for this is overwhelming. Clustering picked out eight separate alignments with travel distances less than 4000 miles, while the Big Ten alignment has a travel distance of more than 6000 miles. Even randomly selected alignments can be expected to outperform the Big Ten alignment with respect to travel distance.

Might the Big Ten alignment have been proposed to maximize attendance? The answer seems to be no. Simple clustering leads quickly to an alignment with a better attendance score than the Big Ten alignment, and the optimal alignment with respect to attendance has a score more than four times that of the Big Ten alignment.

We thus conclude that the new alignment plan was not designed with these aims (minimizing travel distance nor maximizing attendance) in mind. Our study leads to more general conclusions, with regards to the general problem of seperating teams into conferences, as well. Our methodology applies not only to college hockey but other sports and other situations. We have the following suggestions.

When aligning teams into conferences, and searching for an optimal alignment, a combination of objective analysis and subjective personal judgement is recommended. First, clustering should be used to select ten to twenty alignments of interest. These should then be mapped. Next, judgement or optimization theory should be used to shorten this list of alignments to a list of only a select few, optimal alignments. This short list of finalists can then be compared to existing or proposed alignments, and perhaps a clear distinction will become apparent, leading to a single, best alignment.

In closing we make a final comment regarding the speed of calculations. Checking all alignments in a search for an optimal alignment with respect to some attribute such as travel distance or attendance is a time consuming task. Even with our landscape of 21 teams the computations required approximately six days to complete. Clustering on the other hand is a simple procedure that can be carried out in seconds. While it won't necessarily lead to the absolute optimal solution, the evidence that we have presented here indicates that it can be expected to come very close. Since the attributes defining what is optimal are subjective anyway, we conclude that clustering should be the first tool employed in any analysis of the realignment of athletic conferences.

Table 6.1: Clustering resulted in the selection of 11 separate alignments that can be compared based on the criteria of travel distance and attendance.

Alignment Number	Travel Distance (miles)	Attendance Score	Travel Distance (standardized)	Attendance Score (standardized)
1	3079	9.1	-1.19	-0.34
2	3937	16.1	0.49	0.30
3	3168	9.1	-1.02	-0.34
4	3013	11.7	-1.32	-0.10
5	3733	2.0	0.09	-1.01
6	3930	18.8	0.48	0.55
7	4438	9.1	1.49	-0.34
8	4047	8.6	0.71	-0.39
9	4200	3.8	1.01	-0.84
10	3100	11.1	-1.15	-0.16
11	3884	41.8	0.39	2.7

Acknowledgments: Thanks to Gregory Bard for his insight with regards to the clustering method.

References

[1] Big Ten Conference. Men's Ice Hockey To Be Recommended As Official Big Ten Sport Beginning With 2013-14 Academic Year. *Big Ten Athletics News*. General Release, 21 March 2011. Retrieved 8 August 2013 from http://www.bigten.org/genrel/032111aab.html

[2] Horgan, C. Commentary: Conference changes weaken the sport. *USCHO Column*. United States Col- lege Hockey Online, 8 July 2011. Retrieved 8 August 2013 from http://www.uscho.com/2011/07/08/ commentary-conference-changes-weaken-the-sport/

[3] International Hockey Database. Dean Blais. *Hockey Statistics*. Web. Retrieved 8 August 2013 from http://www.hockeydb.com/ihdb/stats/pdisplay.php?pid=8341

[4] NCAA. Champtionship History. *Men's Ice Hockey*. Web. Retrieved 8 August 2013 from http://www.ncaa.com/history/icehockey-men/d1

[5] College Hockey News. Big Ten Officially Announces Hockey Conference. *CHN Staff Report 21* March 2011. Re- trieved 8 August 2013 from http://www.collegehockeynews.com/news/2011/03/21_big_ten_officially_ announces.php

[6] Horgan, C. Commentary: Conference changes weaken the sport. *USCHO Column*. United States Col- lege Hockey Online, 8 July 2011. Retrieved 8 August 2013 from http://www.uscho.com/2011/07/08/ commentary-

conference-changes-weaken-the-sport/
[7] International Hockey Database. Hockey Statistics. Dean Blais Web. Retrieved 8 August 2013 from http://www.hockeydb.com/ihdb/stats/pdisplay.php?pid=8341
[8] Johnson, Richard A. and Wichern, Dean W. (2007) Applied Multivariate Statistical Analysis USA: Pearson.
[9] Lerch, Chris. Commentary: Out of realignment, an opportunity for growth. *USCHO Column* United States College Hockey Online, 8 July 2011. Retrieved 8 August 2013 from http://www.uscho.com/2011/07/08/commentary-out-of-realignment-an-opportunity-for-growth/
[10] Machian, Chris. Uncertainty remains on hockey realignment. *Mavericks Today*. Omaha World-Herald, 15 March 2012. Retrieved 8 August 2013 from http://www.omaha.com/article/20120315/MAVS/711149985
[11] Milewski, Todd. Another view: Thinking of history and the conference realign- ment. *From the Press Box*. United States College Hockey Online, 19 July 2011. Retrieved 8 August 2013 from http://www.uscho.com/from-the-press-box/2011/07/19/another-view-thinking-of-history-and-the-conference-realignment/
[12] As Big Ten prepares to announce plan, CCHA, WCHA commissioners look forward. *USCHO News*. United States College Hockey Online, 16 March 2011. Web. Retrieved 8 August 2013 from http://www.uscho.com/2011/03/16/as-big-ten-prepares-to-announce-plan-ccha-wcha-commissioners-look-forward/
[13] How the 2013–14 conferences look after NCHC goes to 8. *From the Press Box*. United States College Hockey Online, 22 September 2011. Retrieved 8 August 2013 from http://www.uscho.com/from-the-press box/2011/09/22/how-the-2013-14-conferences-look-after-nchc-goes-to-8/
[14] R, The Comprehensive Archive Network. What are R and CRAN? Institute for Statistics and Mathematics of the WU Wien, 2012. Retreived 15 Dec. 2012 from http://cran.us.r-project.org/
[15] United Stats College Hockey Online. Scoreboard. *Box scores*. D–I Men, Season 2011–2012. Retrieved 9 August 2013 from http://www.uscho.com/scoreboard/division-i-

men/2011-2012/week-1/
[16] United States Geological Survey (USGS). Map-It. *Woods Hole Coastal and Marine Science Center* Web. Retreived 8, August 2013 from http://woodshole.er.usgs.gov/mapit/