# Scheduling Soccer:

Designing an Optimal Model for the Great NorthEast Atlantic Conference Soccer Fixtures

Segni & Ye
Dr. Axvig
Dec-18-24

## Introduction

This project focuses on developing an optimized scheduling system for the Great NorthEast Athletic Conference (GNAC), a collegiate athletic association affiliated with NCAA Division III. The Great NorthEast Athletic Conference comprises men's and women's soccer teams from institutions in the northeastern United States.

The primary objective of this scheduling model is to ensure that each team plays matches in a manner that meets various operational and competitive constraints, such as ensuring teams play a balanced number of home and away games, avoiding conflicts where a team is scheduled for multiple games on the same day, accommodating special requests like preferred play dates and so on.

This project employs mathematical programming techniques to formulate and solve the scheduling problem. By leveraging R programming and tools such as lpSolve, lpSolveAPI, and other solvers like CBC from COIN-OR the model integrates various constraints and optimization objectives to generate feasible schedules. These schedules aim to enhance fairness, minimize logistical conflicts, and maximize satisfaction for stakeholders.

While this project focuses on creating an optimal schedule for the current soccer season, the framework is designed with versatility in mind. It can be adapted for future seasons, other divisions within soccer, and even other sports with similar structural requirements. This flexibility ensures that the program remains a valuable tool for addressing diverse scheduling challenges across various contexts.rational challenges.

## Description of the Problem

The Great Northwest Athletic Conference needs a scheduling tool that can be used to schedule their season by year. The conference has 13 Men's teams and 14 women's teams.

There are 13 colleges with both men's teams and women's teams. But, Simmons University (Boston, MA) is a women's university so it only has a women's team. This tool doesn't include the Playoff Schedule, only the regular season.

### When do games happen?

It takes place on Tuesdays, Wednesdays, and Saturdays during the season (for 2025 season: Sat 9/6/25 to Wed 10/29/25). And each team plays one game per date.

**Requests made by the Commissioner**

1. Single Matchups per Date: Each team plays only one game on a given date
2. Balance between Home and Away Games for each team as much as possible
3. Consecutive Games: No teams can play more than three consecutive home or away games
4. Either home or away per season: Teams must play with a specific opponent either home or away once in a season
5. Special Dates => Teams have specific preferences for special dates: homecoming, alumni days, or special events that they want to have the teams at home. These requests are not guaranteed.

# Methodology

This section details the tools, constraints, optimization objectives, and mathematical formulation employed in developing the GNAC scheduling model.

## Tools and Software

To build and solve the scheduling model, the following tools and software were used:

- **R Programming**: The primary language for data manipulation, model formulation, and solution visualization.
- **Libraries**:
  - lpSolve and lpSolveAPI: For linear programming and optimization.
  - readxl: To read team and schedule data from Excel files.
  - lubridate: For date handling and manipulation.
- **Excel**: For input data storage, including team rosters, play dates, and special requests.
- **COIN-OR CBC Solver**: To solve large-scale mixed-integer programming problems.

## Problem Constraints and Optimization Objectives

The scheduling model incorporates multiple constraints and optimization objectives to ensure fairness and feasibility. Each team plays an approximately equal number of home and away games throughout the season, with only one match (home, away, or a break playdate) scheduled per date. Teams face each specific opponent only once per season, either at home or away. To maintain balance, no team plays more than three consecutive home or away games. Men's teams are limited to 12 games per season, while women's teams play 13 games. Additionally, teams cannot be scheduled to play on both Tuesday and Wednesday within the same week.

**Special Requests (Objective Function)**:

Accommodates preferences for home games on specific dates for events such as homecoming or alumni days, though these are not guaranteed.

| Teams | Team Name | Special Dates |
|---|---|---|
| AMC | Albertus Magnus College | TBD (Homecoming) |
| CSC | Colby-Sawyer College | October 18, 2025 (Homecoming), October 4, 2025 (Athletics Open House) |
| DC | Dean College | October 4, 2025 (Homecoming) |
| EC | Elms College | TBD (Homecoming) |
| ENC | Emmanuel College | October 4, 2025 (Homecoming) |
| LU | Lasell University | October 17-19, 2025 (Friends and Family Weekend) |
| MC | Mitchell College | October 26, 2025 (Fall Weekend) |
| NEC | New England College | October 4, 2025 (Fall Festival) |
| NU | Norwich University | September 19-20, 2025 (Homecoming), October 3-4, 2025 (Parent's Weekend) |
| RU | Regis College | September 27, 2025 (Homecoming/RegisFest) |
| RRU | Rivier University | September 27, 2025 (Parents and Alumni Weekend) |
| SJM | Saint Joseph's College of Maine | October 17-19, 2025 (Homecoming) |
| SMU | Simmons University | October 4, 2025 (Alumnae Day) |
| USJ | Saint Joseph University Connecticut | October 4, 2025 (Family Weekend/Homecoming) |

# Mathematical Model

## Model Creation

Let $T$ be the names of each of the teams. Note here that I'm treating Simmons University ($SMU$) as a bye when it comes to a men's team "playing" SCU. For notational convenience, define $T_* := T\backslash\{SMU\}$.

Let $G = \{Men, Women, M\}$ be the set of sexes for a team, let $D$ be the set of dates in a season and let $D_w$ be all dates that occur on a weekday (play only happens on Tuesdays, Wednesdays and Saturdays). We partition $D$, the playdates for the season. For each $t \in T$ and each $g \in G$, we define $T_{1,t,g}$ to be the specific teams team $t$ of sex $g$ needs to play at some point in the season (note that for men's teams this set will not include $SMU$, we're tracking only real opponents). $T_{2,t,g}$ is defined similarly. For any $d \in D$, define $w(d)$ to be the (numeric) day of the week that $d$ lands on (Wednesdays are 4's, Saturday's are 7's). This will serve as a redundant subscript in our variables to help out writing certain constraints.

$S_d$ be the set of special dates requested by the colleges in the GNAC conference.

## Decision Variables

For all $t_1 \in T_*$, $d \in D_1$, and all $t_2 \in T_{1,t,Men}$ define

$$x_{t_1,t_2,d,w(d),Men} = \{1 \ \ if\ men's\ team\ t_1\ hosts\ t_2\ on\ d,\ 0\ \ otherwise$$

For all $t_1 \in T$, $d \in D_1$, and all $t_2 \in T_{1,t,Women}$ define

$$x_{t_1,t_2,d,w(d),Women} = \{\ 1\ \ if\ women's\ team\ t_1\ hosts\ t_2\ on\ d,\ 0\ otherwise$$

For all $t \ \in T_*$, $d \ \in D$ , define

$$b_{t,d,Men} = \{1 \ \ if\ men's\ team\ t\ \ doesn't\ play\ on\ d,\ 0\ \ otherwise$$

For all $t \ \in T$ , $d \ \in D$ , define

$$b_{t,d,Women} = \{1 \ \ if\ women's\ team\ t\ \ doesn't\ play\ on\ d,\ 0\ \ otherwise$$

## Constraints

1. **Balancing Home & away games**
    a. For Women's Teams

    For each women's team, home and away games are balanced. It is done by making the total number of home or away games to be greater than or equal to(>=) 5 games. Since the total games for each women's team is 13, each team will have (>=5) and (<=7) games either home or away.

    For all $t_1 \in T$:

    Home games: $\sum_{d \in D} \sum_{t_2 \in T \setminus t_1, women} x_{t1,t2,d,w(d),Women} >= 5$

    Away games: $\sum_{d \in D} \sum_{t_2 \in T \setminus t_1, women} x_{t2,t1,d,w(d),Women} >= 5$

    b. For men's teams

    The same reasoning as above, but as men's total games for each team is 12 games, each men's team will have (>=5) and (<=6) either home or away per season.

For all $t_1 \in T_m$:

Home games: $\sum_{d \in D} \sum_{t_2 \in T \setminus t_1, men} x_{t1,t2,d,w(d),men} >= 5$

Away games: $\sum_{d \in D} \sum_{t_2 \in T \setminus t_1, men} x_{t2,t1,d,w(d),men} >= 5$

2. **Each team only play once per available playdate either home or away**
   a. For women's teams

      For all $t_1 \in T$ and $d \in D$:

$$\sum_{t_2 \in T \setminus t_1, women} x_{t1,t2,d,w(d),Women} + \sum_{t_2 \in T \setminus t_1, women} x_{t2,t1,d,w(d),Women} + b_{t_1,d,Women} = 1$$

   b. For men's teams

      For all $t_1 \in T_m$ and $d \in D$:

$$\sum_{t_2 \in T \setminus t_1, men} x_{t1,t2,d,w(d),men} + \sum_{t_2 \in T \setminus t_1, men} x_{t2,t1,d,w(d),men} + b_{t_1,d,Men} = 1$$

3. **A team should play only once with a specific opponent either home or away in a season.**
   a. For women's teams

      For all $t_1 \in T \ and\ t_2 \in T_{t1,women}$:

$$\sum_{d \in D} x_{t1,t2,d,w(d),women} + \sum_{d \in D} x_{t2,t1,d,w(d),women} = 1$$

   b. For men's teams

      For all $t_1 \in T_m \ and\ t_2 \in T_{t1,men}$:

$$\sum_{d \in D} x_{t1,t2,d,w(d),men} + \sum_{d \in D} x_{t2,t1,d,w(d),men} = 1$$

4. **No more than three consecutive home games per season**
   a. For women's teams

      For all $t_1 \in T$ :

$$\sum_{d \in \{d_k, d_{k+1}, d_{k+2}, d_{k+3}\}} \sum_{t_2 \in T \setminus t_1, women} x_{t1,t2,d,w(d),women} <= 3$$

b. For men's teams

For all $t_1 \in T_m$ :

$$\sum_{d \in \{d_k, d_{k+1}, d_{k+2}, d_{k+3}, d_{k+4}\}} \sum_{t_2 \in T \setminus t_1, men} x_{t1,t2,d,w(d),men} <= 3$$

5. **No more than three consecutive away games per season**
   a. For women's teams

   For all $t_1 \in T$ :

$$\sum_{d \in \{d_k, d_{k+1}, d_{k+2}, d_{k+3}\}} \sum_{t_2 \in T \setminus t_1, men} x_{t2,t1,d,w(d),women} <= 3$$

   b. For men's teams

   For all $t_1 \in T_m$ :

$$\sum_{d \in \{d_k, d_{k+1}, d_{k+2}, d_{k+3}, d_{k+4}\}} \sum_{t_2 \in T \setminus t_1, men} x_{t2,t1,d,w(d),men} <= 3$$

6. **Men's teams only play 12 games per season either home or away.**

   For all $t_1 \in T_m$ :

$$\sum_{d \in D} \sum_{t_2 \in T \setminus t_1, men} x_{t1,t2,d,w(d),men} + x_{t2,t1,d,w(d),men} = 12$$

7. **Women's teams only play 13 games per season either home or away.**

   For all $t_1 \in T$ :

$$\sum_{d \in D} \sum_{t_2 \in T \setminus t_1, women} x_{t1,t2,d,w(d),women} + x_{t2,t1,d,w(d),women} = 12$$

## Objective Function

The objective is to maximize the special date satisfaction at home games as per requested. The satisfaction is 5 for either man or women playing at home on the special date and zero if they don't.

**Model Implementation**

The model implementation involved several key steps, beginning with data initialization, where team and date data were imported from Excel, and decision variables were generated for all possible matchups. Next, a sparse constraint matrix was constructed, with rows representing the constraints and columns representing the decision variables. The solver was then executed using lpSolve to optimize the binary linear programming problem and generate a feasible schedule. Finally, the generated schedule was validated against the constraints, and the results were visualized to ensure the schedule's practical usability.

# Results, Recommendations, and Instructions

## Instructions for Use

1. **Set Up Your Environment**

- **Install Required R Libraries**: Ensure you have the required R libraries installed, such as:

```
install.packages("lpSolveAPI")
install.packages("lubridate")
```

- **Install CBC Solver**: Download and install the CBC solver if not already installed. You can find it here. Place the binary file in a known directory and ensure it is executable.
- **Define Working Directory**: Update the script with the correct paths for the working directory and output files. Modify these lines in the script to match your system:

```
# Define paths for the main project file and CBC solver directory
mainfile = "path/to/your/directory"
coinRDir = "path/to/your/directory"
coinRSoltuinDir = "path/to/your/directory/LPSolution.txt"



# Read all sheet names from the Excel file
excel_sheets("GNAC_Soccer.xlsx")
```

```
# Read men's teams, women's teams, and play dates from the Excel file
MenTeams <-read_excel("GNAC_Soccer.xlsx",
            sheet = "TeamsMens")
WomenTeams <-read_excel("GNAC_Soccer.xlsx",
            sheet = "TeamsWomens")
Dates <-read_excel("GNAC_Soccer.xlsx",
            sheet = "2025Dates")
```

2. **Run the LP Problem Definition Section**
   - Execute the section of the script that defines and writes the LP problem in MPS format.
   - Save the MPS file to the specified location for solving:

     ```
     write.lp(LP, 'problem.mps', type = 'mps')
     ```

3. **Solve the LP Problem**
   - Use the CBC solver to solve the problem by running the following command in the terminal or system shell:

     ```
     cbc problem.mps maxN 10000000 solve solution LPSolution.txt exit
     ```

   - Ensure the solution is saved as `LPSolution.txt` in the expected directory.

4. **Extract and Parse the Solution**
   - Run the extraction and parsing section to generate the `fullSolutionVector` and `solutionVector`.
   - Exporting Files->
     womensCSV<-"path/to/your/result/directory"

5. **Generate Schedules**
   - Populate the schedules for men's and women's teams using the extracted solution vector.

   - This will create matrices `scheduleMen` and `scheduleWomen`, containing the match details.

6. **Export Schedules**
   - Save the schedules as CSV files for further use:

     ```
     write.csv(scheduleMen, file = mensCSV, row.names = TRUE)
     write.csv(scheduleWomen, file = womensCSV, row.names = TRUE)
     ```

7. **Validate and Analyze Results**

- Open and validate the CSV files (mens_schedule.csv and womens_schedule.csv) to confirm the schedules are correct.
- Use the View(scheduleMen) or View(scheduleWomen) commands in R for quick visualization.

**WorkFlow Overview**

The workflow for the scheduling model begins with input preparation, where teams, dates, and constraints are defined. This is followed by model definition, which sets up the objective function, constraints, and linear programming (LP) problem. The problem-solving phase involves exporting the model to MPS format, running the CBC solver, and parsing the solution. Finally, schedules are generated and saved for distribution.

For example, in a soccer tournament, organizers can list participating teams, define matchday constraints, and use the model to create optimal schedules that balance home and away games while adhering to the constraints.

**Results**

The scheduling model showcases several strengths, including its reusability with minimal input via Excel, enabling the creation of annual schedules in a short amount of time. It can generate readable schedules in Excel or CSV formats, making them accessible to users without requiring knowledge of R or RStudio. The model is adaptable to any number of teams, dates, or special date requests, making it suitable for soccer conferences, though it has not yet been tested with large datasets. Importantly, it successfully met all requested dates from colleges and fulfilled the requirements set by Joe Walsh, the GNAC commissioner.

However, the model has some weaknesses. Installing the CBC/COIN-OR solver can be challenging. We had to separate Men's and Women's schedule files due to differences in schedules between them.

**Assumptions**

The model operates under several assumptions, including that teams can adjust their schedules to play on all the provided playdates and that each team will play at least five home and five away games, though the balance between home and away games is not strictly enforced. The season is assumed to begin on the second Saturday and end on the last Wednesday of the provided playdates, with no distinction made if men's and women's teams play on the same day. All requested dates are considered equally important, and the model assumes a maximum of one break for men's teams and two for women's teams.

**Validity of our Solution**

The solution demonstrates strong validity, ensuring that both men's and women's teams rested for only two playdates. However, due to the odd number of men's teams, each had one one-playdate break. The input information proved useful in constructing the model, though limitations exist. For instance, the model would fail with extended playdate availability or if the constraint for consecutive home or away games were violated.

**Recommendations**

To enhance the scheduling process, the following recommendations are proposed: the GNAC should adopt the model as a standard tool for consistent and optimized yearly scheduling. Establishing feedback loops with teams, coaches, and administrators after each season can highlight areas for improvement. The model's versatility suggests it could be adapted for other sports like basketball or volleyball. Finally, incorporating travel minimization and cost considerations as optimization objectives could improve the model further, particularly for schools with tighter budgets.

# Conclusion

This project successfully developed an optimized scheduling tool for GNAC soccer fixtures, meeting critical operational and competitive constraints. The model's adaptability ensures its utility beyond the current season.

The scheduling tool not only simplifies fixture creation but also enhances fairness and stakeholder(commissioner) satisfaction. Its flexibility enables it to accommodate various constraints, such as balancing home and away games, minimizing consecutive matchups, and addressing special requests. These features make it a valuable resource for GNAC and a template for similar scheduling issues elsewhere.

In addition, this project underscores the importance of optimization techniques in sports management. By addressing intricate constraints and incorporating stakeholder input, the model achieves a fine balance between operational efficiency and competitive integrity. Looking forward, this framework can inspire innovations in scheduling for other collegiate conferences, fostering better collaboration and fairness across sports organizations.

# Works Cited


- R Documentation: lpSolve, lpSolveAPI, readxl, lubridate.
- COIN-OR CBC Solver Documentation.
- GNAC Official Website (for team and scheduling information).