

CONSTRUINDO UM CALENDÁRIO MAIS JUSTO PARA O CAMPEONATO BRASILEIRO DE FUTEBOL CONSIDERANDO DOIS GRUPOS DE FORÇA

Marcos Costa Roboredo¹

Luiz Aizemberg²

João Carlos C. B. Soares de Mello³

Artur Alves Pessoa⁴

RESUMO

O campeonato brasileiro de futebol é um clássico torneio com dois turnos espelhados. Este campeonato é composto de divisões (séries), onde as duas mais fortes são denominadas A e B, nesta ordem. Atualmente, os quatro primeiros colocados times da Série A e o vencedor da Copa do Brasil disputam a Copa Libertadores no próximo ano enquanto os quatro primeiros colocados times da Série B disputam a Série A no próximo ano. Um efeito *carry-over* ocorre no calendário quando um time enfrenta dois times fortes (fracos) consecutivamente. Neste trabalho, consideramos um problema de programação inteira que constrói a tabela de jogos mais justa para o campeonato brasileiro de futebol minimizando o número de efeitos *carry-over*. Mostramos ainda evidências de que times da Serie B na temporada anterior e disputando a Serie A na temporada atual formam um grupo fraco, enquanto times disputando a Copa Libertadores jogando a Serie A na atual temporada formam um grupo forte.

Palavras-chave: efeitos *carry-over*, calendários esportivos, torneios espelhados com dois turnos.

1.Introdução

Em um torneio simples com um turno, no inglês *Single Round Robin tournament*, com um número n par de times, cada time enfrenta outra equipe exatamente uma vez. O torneio é dividido em $n - 1$ rodadas, onde cada time joga uma vez por rodada em casa (em seu próprio estádio) ou fora (estádio do oponente). Já um torneio com dois turnos espelhados, no inglês *Mirrored Double Round Robin Tournament*, é um torneio composto

¹ Universidade Federal Fluminense. Email: mcr.marcos@yahoo.com.br

² Universidade Federal Fluminense. Email: luizaizemberg@gmail.com

³ Universidade Federal Fluminense. Email: jcamello@yahoo.com.br

⁴ Universidade Federal Fluminense. Email: artur@producao.uff.br

por dois torneios simples com um turno, onde o segundo é muito semelhante ao primeiro com uma única diferença: o mando de campo de cada partida é inverso. Vamos introduzir a seguinte notação clássica sobre torneios esportivos com n times. “Padrão” é uma sequência 0-1 indicando se um time joga em casa (1) ou fora (0). O chamado *break* ocorre no calendário rodada $r + 1$ se um time joga em casa (fora), tanto na rodada r quanto na rodada $r + 1$. Para exemplificar, consideremos o padrão 0 1 0 0 1 0 1 1 do time A. Isto indica que A joga em casa (fora) nas rodadas 2,5,7 e 8 (1,3,4 e 6) e possui dois *breaks*, nas rodadas 4 e 8. Qualquer conjunto composto de n padrões é chamado simplesmente de “conjunto de padrões”. Um padrão é dito complementar a outro caso os mandos de campo entre estes sejam inversos.

Na literatura, podemos encontrar diversos trabalhos relacionados a construção de torneios (Rasmussen & Trick, 2008). A maioria destes resolvem o problema usando o procedimento proposto por Nemhauser e Trick (1998), que consiste em construir um torneio simples através de três fases: a primeira gera uma grande quantidade de conjunto de padrões, a segunda tenta construir um calendário viável com os conjuntos de padrões gerados enquanto a última fase atribui cada equipe a um padrão e garante, caso haja, restrições locais. Algumas restrições globais para construção de calendários esportivos são apresentadas em Henz *et al* (2004), enquanto restrições específicas para o campeonato brasileiro de futebol são tratadas em Ribeiro e Urrutia (2009). A construção de calendários para as ligas de futebol austríaca e alemã, chilena, dinamarquesa e holandesa são tratadas, respectivamente, em Bartsch *et al* (2006), Duran *et al* (2007), Rasmussen (2008) e Schreuder (1992).

Com objetivo de obter calendários mais justos para os torneios, a questão mais importante é evitar *breaks* (Miyashiro *et al*, 2003) uma vez que se um time possui jogos consecutivos em casa (fora) então pode gerar vantagens ou desvantagens para este. Além disso, outros objetivos têm sido propostos associados a grupos de força no campeonato. Della Croce e Oliveri (2006), por exemplo, consideraram um grupo de times cabeças de chave, onde jogos entre os times pertencentes a este grupo não ocorrem nas primeiras (últimas) três rodadas. Já Ribeiro e Urrutia (2006) consideraram um grupo de times de elite, onde times não enfrentam times deste grupo cinco vezes consecutivas. Russell (1980) propôs evitar os chamados efeitos *carry-over* nos calendários.

Um efeito *carry-over* ocorre quando um time enfrenta duas vezes consecutivas dois times fortes ou dois times fracos. Briskorn (2009) e Briskorn e Knust (2010) consideraram partições de grupos de fortes e fracos, mostrando se existe um calendário sem efeito *carry-over* para diversos valores de número de times e diversos tamanhos para os grupos de força. Goossens e Spieksma (2011) fazem um estudo sobre efeitos *carry-over* em toda a Europa e de acordo com Goossens e Spieksma (2009), a liga de futebol belga foi a primeira a considerar a relevância do

balanceamento dos efeitos carry-over no calendário.

Neste trabalho, nós tratamos o campeonato brasileiro de futebol, que é um clássico torneio espelhado com dois turnos, um número par de times e algumas restrições locais. Este campeonato é composto de divisões (séries), onde as mais fortes são as 1ª e 2ª divisões (séries A e B), nesta ordem. Atualmente, em um determinado ano, os quatro primeiro colocados da série A e o campeão da Copa do Brasil recebem o direito de disputar a importante Copa Libertadores da América no próximo ano enquanto os quatro primeiro colocados da série B recebem o direito de disputar a série A no próximo ano. Neste contexto, surgem duas hipóteses intuitivas. Hipótese 1 (H1) diz que times vindos da série B em um ano precedente e que disputam a série A no ano corrente tem grandes chances de terminar o campeonato na metade de baixo da tabela de classificação. Hipótese 2 (H2) diz que times que jogam a Copa Libertadores da América e disputam a série A no ano corrente tem grandes chances de terminar a série A na metade de cima da tabela de classificação.

Para ambos os casos, a hipótese nula (H0) diz que as posições dos times na tabela de classificação são independentes se eles jogaram a série B no ano precedente ou estão disputando a Copa Libertadores da América no ano corrente. Primeiramente, nós mostramos que H1 e H2 são verdadeiras, rejeitando H0 baseados no número de times que satisfazem as hipóteses em 19 campeonatos brasileiros passados. Nós acreditamos que isto pode ser usado como argumento em favor de adicionar restrições para evitar efeitos *carry-over* ao modelo para geração da tabela do campeonato. Neste contexto, nós construímos um calendário para série A do campeonato brasileiro de futebol através de um modelo de Programação Inteira (PI) similar ao proposto por Della Croce and Oliveri (2006) e, com objetivo de um calendário final mais justo, nós garantimos os seguintes requerimentos: o número final de efeitos *carry-over* é mínimo, cada time tem no máximo 4 *breaks* e não existem *breaks* consecutivos.

Este trabalho está dividido como segue. Seção 2 mostra os resultados do teste de hipótese associado aos grupos de força propostos. Seção 3 detalha as fases de construção do calendário. Seção 4 mostra nossos resultados computacionais, comparando os calendários gerados pelo nosso procedimento com os calendários oficiais de 2006 até 2010 com relação ao número de efeitos *carry-over*. Finalmente, conclusões e perspectivas de trabalhos futuros são apresentadas na seção 5.

2. Teste de Hipótese

Um teste de hipótese foi aplicado usando dados reais do campeonato brasileiro de futebol de 1990 à 2009 com intuito de analisar as seguintes hipóteses:

- **H1:** Times vindos da série B em um ano precedente e que

disputam a série A no ano corrente tem grandes chances de terminar o campeonato na metade de baixo da tabela de classificação.

- **H2:** Times que jogam a Copa Libertadores e disputam a série A no ano corrente tem grandes chances de terminar a série A na metade de cima da tabela de classificação.

A hipótese nula (H0) para as duas hipóteses prévias é a que segue:

- **H0:** As posições dos times na tabela de classificação são independentes se eles jogaram a série B no ano precedente ou estão disputando a Copa Libertadores da América no ano corrente.

Primeiramente, contamos os times que se enquadram em cada uma das hipóteses (H1 e H2), dividindo pelo número total de times fracos (fortes) de acordo com as hipóteses. Como esperávamos, os resultados estão longe de 50%: 65,51% para H1 e 68,57% para H2. Assim, calculamos o p-valor de H0 para os dados tanto para H1 quanto H2, obtendo 0,0045 para H1 e 0,0004 para H2. Baseados nestes valores, temos boas evidências para rejeitar H0.

3. O procedimento

Para gerar o calendário final, nós usamos um procedimento de quatro fases que é uma adaptação do procedimento proposto por Nemhauser e Trick (1998).

Fase 1: Geração randômica de um grande número de conjuntos de padrões, limitando o total de *breaks* de cada padrão.

Fase 2: Para cada conjunto de padrões gerado na fase anterior, testar se é possível construir um calendário viável com este através do modelo de programação linear proposto por Briskorn (2008).

Fase 3: Para cada conjunto de padrões que passaram no teste da fase anterior, construir um calendário viável com este através de Programação Inteira.

Fase 4: Para cada calendário construído na fase anterior, atribuir times aos padrões de modo a minimizar o número de efeitos *carry-over*.

Agora, nós detalhamos cada uma das 4 fases. Além disso, na seção

3.5, ilustramos todas estas fases através de um simples exemplo.

3.1 Fase 1: Gerando conjunto de padrões.

Nessa fase, nós geramos aleatoriamente uma grande quantidade de conjuntos de padrões (em torno de 200), garantindo três requerimentos:

- Cada padrão é composto por no máximo quatro *breaks*.
- Sem *breaks* consecutivos.
- Os conjuntos de padrões gerados são compostos por pares de padrões complementares.
- A metade dos jogos de cada padrão é em casa e a outra metade, fora.

3.2 Fase 2: Testando a viabilidade de cada conjunto de padrões.

Nessa fase, nós testamos se é possível construir um calendário para cada conjunto de padrões gerados na fase 1 usando o modelo de programação linear proposto por Briskorn (2008). Aqui, dois aspectos computacionais são importantes: o tempo computacional para testar cada conjunto de padrões é desprezível e em torno de 50% do total de conjuntos de padrões gerados na fase 1 são viáveis.

3.3 Fase 3: Gerando calendários.

Nessa fase, para cada conjunto de padrões validado na fase 2, nós escalonamos um calendário através do modelo de programação inteira sem função objetivo, mostrado a seguir. Seja x_{ijr} variável binária que indica quando o padrão i enfrenta o padrão j na rodada r ($i > j$). Seja Υ_{ir} uma constante que indica quando o padrão i joga em casa ($\Upsilon_{ir} = 1$) ou fora de casa ($\Upsilon_{ir} = 0$) na rodada r . A seguir, temos as restrições do modelo:

$$\sum_j x_{ijr} = 1, \quad \forall i, r \quad (1)$$

$$\sum_r x_{ijr} = 1, \quad \forall i, j \quad (2)$$

$$x_{ijr} \leq |\Upsilon_{ir} - \Upsilon_{jr}|, \quad \forall i, j, r \quad (3)$$

$$x_{ijr} \in \{0, 1\}, \quad \forall i, j, r \quad (4)$$

As restrições (1) garantem para cada padrão i , para cada rodada r , i joga contra exatamente um time. As restrições (2) garantem que, para cada par de padrões i e j , existe uma rodada onde i e j se enfrentam, enquanto (3)

garantem que um jogo entre padrões i e j na rodada r só pode ocorrer se o mando de campo destes é inverso. É válido ressaltar que são necessários poucos segundos para a execução do modelo prévio.

3.4 Fase 4: Atribuir times aos padrões.

Nesta fase, nós apresentamos um modelo de programação inteira onde, para cada calendário da fase 3, cada padrão é atribuído a exatamente um time minimizando o número de efeitos de *carry-over* total do calendário e garantindo algumas restrições básicas para torneios esportivos. O modelo é o que segue. Seja x_{ij} variável binária indicando se o padrão i é atribuído ao time j . Seja y_{ijr} (z_{ijr}) variável binária indicando se um efeito *carry-over* ocorre quando o time atribuído ao padrão i joga consecutivamente contra o par de times fortes (fracos) j nas rodadas r e $r+1$. Nós definimos ainda a seguinte notação. R denota o conjunto de padrões, T denota o conjunto de times, P_l denota o conjunto de padrões escalonado na fase 3, $O_{|P_l| \times |R|}$ é uma matriz onde cada célula O_{ir} indica o padrão oponente do padrão i na rodada r de acordo com o calendário da fase 3, E é o conjunto de pares de times que usam o mesmo estádio para jogos em casa, C é o conjunto de pares de padrões complementares, S (W) é o conjunto composto por pares de times fortes (fracos). O modelo completo segue abaixo:

$$\text{Minimizar} \quad \sum_{i \in P} \sum_{j \in S} \sum_{r \in R} y_{ijr} + \sum_{i \in P} \sum_{j \in W} \sum_{r \in R} z_{ijr} \quad (5)$$

Sujeito a:

$$\sum_{j \in T} x_{ij} = 1, \quad \forall i \in P_1 \quad (6)$$

$$\sum_{i \in P} x_{ij} = 1, \quad \forall j \in T \quad (7)$$

$$x_{c_1 j_1} - x_{c_2 j_2} = 0, \quad \forall (c_1, c_2) \in C, \forall (j_1, j_2) \in E \quad (8)$$

$$x_{O_{ir} j_1} + x_{O_{r+1} j_2} - y_{ijr} \leq 1, \quad \forall i \in P_1, \forall j = (j_1, j_2) \in S, \quad \forall r \in \{1, \dots, |R| - 1\} \quad (9)$$

$$x_{O_{|R|} j_1} + x_{O_{1} j_2} - y_{ij|R|} \leq 1, \quad \forall i \in P_1, \forall j = (j_1, j_2) \in S \quad (10)$$

$$x_{O_{ir} j_1} + x_{O_{r+1} j_2} - z_{ijr} \leq 1, \quad \forall i \in P_1, \forall j = (j_1, j_2) \in W, \quad \forall r \in \{1, \dots, |R| - 1\} \quad (11)$$

$$x_{O_{|R|} j_1} + x_{O_{1} j_2} - z_{ij|R|} \leq 1, \quad \forall i \in P_1, \forall j = (j_1, j_2) \in W \quad (12)$$

A função objetivo (5) minimiza o número total de efeitos *carry-over*. As restrições (6) garantem que cada padrão é atribuído a exatamente um time. As restrições (7) garantem que cada time j é atribuído a exatamente

um padrão. As restrições (8) garantem que os times que usam o mesmo estádio para jogos em casa são atribuídos a padrões complementares enquanto as restrições (9), (10), (11) e (12) calculam a quantidade de efeitos *carry-over*. O tempo de processamento do modelo computacional da fase 4 depende de n , $|S|$ e $|W|$. O pior caso testado é considerando $n = 20$, $|S| = 6$ e $|W| = 4$. Nesse caso, o tempo médio para as primeiras 10 soluções é de cinco minutos (PC Core i5 2,27GHz, 4GB Ram, Windows 7, Visual Basic).

3.5 Exemplo ilustrativo.

Para ilustrar cada fase do procedimento, nós apresentamos um exemplo simples. Ele possui somente $n = 4$ times denotados por A, B, C e D. Fase 1 gera uma certa quantidade de padrões. $P_1: 101$, $P_2: 010$, $P_3: 100$ e $P_4: 011$ representam um exemplo de conjunto de padrões para o problema ilustrativo. Fase 2 testa cada conjunto de padrões da fase 1 quanto à viabilidade. O conjunto de padrões mostrado como exemplo foi testado e é viável. A próxima fase constrói calendários com os conjuntos de padrões viáveis encontrados na fase 2. Um possível calendário para o conjunto de padrões dado como exemplo está na tabela 1. A fase 4 atribui padrões a times. Por exemplo, $P_1 \rightarrow B$, $P_2 \rightarrow D$, $P_3 \rightarrow A$ e $P_4 \rightarrow C$. O escalonamento final para o exemplo ilustrativo está na tabela 2.

Tabela 1. Um possível calendário para o conjunto de padrões do exemplo ilustrativo

Rodada 1	Rodada 2	Rodada 3
$P_1 - P_2$	$P_1 - P_4$	$P_1 - P_3$
$P_3 - P_4$	$P_2 - P_3$	$P_2 - P_4$

Tabela 2. Calendário final para o exemplo ilustrativo

Rodada 1	Rodada 2	Rodada 3
B - D	B - C	B - A
A - C	D - A	D - C

4. Resultados computacionais

Nós testamos nosso algoritmo usando dados do campeonato brasileiro de futebol entre 2006 e 2010. Para todos os anos, somente dois times usaram o mesmo estádio para jogos em casa: Flamengo e Fluminense. O resumo dos resultados obtidos está mostrado na tabela 3. Coluna 1 indica os anos considerados. Coluna 2, 3 e 4 mostram o ano corrente, o número de times, o número de times pertencentes ao grupo forte e o número de times pertencentes ao grupo fraco, respectivamente. O número de efeitos *carry-over* no calendário oficial é mostrado respectivamente nas colunas 5, 6 e 7. Colunas 8, 9 e 10 mostram a quantidade média de efeitos *carry-over* dos 10 primeiros calendários encontrados pelo algoritmo proposto.

Esses resultados mostram que, para todos os anos, o número de efeitos *carry-over* na solução proposta é significativamente menor comparado ao calendário oficial. Apesar de nós não estarmos considerando as restrições locais do campeonato brasileiro, a grande redução obtida no número de efeitos *carry-over* sugere que mesmo considerando todas as restrições do calendário oficial, o efeito *carry-over* poderia ser substancialmente reduzido.

5. Conclusões e trabalhos futuros

Esse artigo propôs escalonar o campeonato brasileiro de futebol de maneira mais justa considerando uma pequena quantidade de *breaks* e minimizando a quantidade total de efeitos *carry-over* associados a dois grupos de força. Nós justificamos o uso destes grupos através de análise estatística dos campeonatos passados. O método utilizado para gerar os calendários finais é uma simples extensão do método de decomposição mais utilizado na literatura para o problema. Os calendários gerados pelo nosso método possuem um número significativamente menor de efeitos *carry-over* quando comparados aos calendários oficiais para diversos anos. Além disso, é de fácil incorporação a função objetivo e as restrições propostas na fase 4 a qualquer modelo de programação inteira para escalonar calendários esportivos também baseados no método da decomposição. Como trabalhos futuros, planejamos incluir restrições locais do campeonato brasileiro para obter calendários mais próximos à realidade.

Tabela 3. Resultados computacionais

Ano	N	S	W	<i>carry-over</i> (oficial)			<i>carry-over</i> (este artigo)		
				Forte	Fraço	Total	Forte	Fraço	Total
2010	20	5	4	12	14	26	6,7	2,1	8,8
2009	20	5	4	27	18	45	6,7	2,1	8,8
2008	20	5	4	19	13	32	6,7	2,1	8,8
2007	20	6	4	31	6	37	13,3	2,2	15,5
2006	20	5	2	17	0	17	6,5	0	6,5

Referências

- (1) Bartsch, T., Drexl, A., Kroger S. (2006) Scheduling the professional soccer leagues of Austria and Germany. *Computer and Operations Research*, 33(7) 1907–1937.
- (2) Briskorn D. (2008) Feasibility of home-away-pattern sets for round robin tournaments. *Operations Research Letters* 36, 283–284.
- (3) Briskorn D. (2009) Combinatorial properties of strength groups in round robin tournaments. *European Journal of Operational Research* 192, 744–754.

- (4) Briskorn D. and Knust S. (2010) Constructing fair sports league schedules with regard to strength groups. *Discrete Applied Mathematics* 158, 123–135.
- (5) Della Croce F. and Oliveri D. (2006) Scheduling the Italian football league: An ILP-based approach. *Computers & Operations Research* 33, 1963–1974.
- (6) Duran, G., Guajardo, M., Miranda, J., Sauré, D., Souyris, S., Weintraub, A. (2007) Scheduling the Chilean soccer league by integer programming. *Interfaces* 37(6) 539–552.
- (7) Goossens, D. e Spieksma, F. (2009) Scheduling the Belgian Soccer League. Vol. 39, No. 2, 109–118, issn 0092-2102 eissn 1526-551X 09 3902 0109.
- (8) Goossens, D., Spieksma, F. (2011) Soccer schedules in Europe: an overview. *Journal of Scheduling*, Issn: 1094-6136, 1-11.
- (9) Henz, M., Müller, T., Thiel, S. (2004) Global constraints for round robin tournament scheduling. *European Journal of Operational Research*, 153 (1), 92-101.
- (10) Miyashiro, R., Iwasaki, H., Matsui, T. (2003) Characterizing feasible pattern sets with a minimum number of breaks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2740, 78-99.
- (11) Nemhauser, G. and Trick, M. (1998) Scheduling a major college basketball conference. *Operations Research* 46, 1–8.
- (12) Rasmussen, R.V., Trick, M.A. (2008) Round robin scheduling - a survey. *European Journal of Operational Research*, 188 (3), 617-636.
- (13) Rasmussen, R. V. (2008) Scheduling a triple round robin tournament for the best Danish soccer league. *European Journal of Operational Research*, 185(2) 795–810.
- (14) Ribeiro, C. C. and Urrutia, S. (2006) Scheduling the Brazilian soccer tournament with fairness and broadcast objectives, in *Proceedings of the 6th international conference on Practice and theory of automated timetabling VI*, Springer-Verlag, 147–157.
- (15) Ribeiro, C. C. and Urrutia, S. (2009) Bicriteria integer programming approach for scheduling the Brazilian national soccer tournament. *Proceedings of the 3rd International Conference on Management Science and Engineering Management*, 55-58.
- (16) Russell K. (1980) Balancing carry-over effects in round robin tournaments. *Biometrika*, 67, 127.
- (17) Schreuder, J. A. M. (1992) Combinatorial aspects of construction of competition Dutch professional football leagues. *Discrete Appl. Math.*, 35(3) 301–312.