



Análise experimental da complexidade do problema da minimização da consanguinidade em sistemas de acasalamento

Éric Dias da Silva Rosso¹, Ana Paula Lüdtke Ferreira¹

¹Programa de Pós-graduação em Computação Aplicada
Universidade Federal do Pampa, Campus Bagé

{ericrosso.aluno, anaferreira}@unipampa.edu.br

Resumo. *O melhoramento animal depende de sistemas de acasalamento para produzir a melhor próxima geração do rebanho. Além do ganho genético e econômico esperado, a consanguinidade do rebanho deve ser mantida baixa, para evitar problemas gerados por depressão endogâmica. Este trabalho objetiva iniciar o estudo da complexidade do problema da minimização da consanguinidade usando uma combinação de programação linear com branch-and-bound, para investigar o tempo de execução médio da solução do problema. Os resultados apontam que o problema parece ser computacionalmente intratável.*

Abstract. *Animal breeding relies on mating systems to produce the best possible herd's next generation. Besides expected genetic and economic gain, keeping low inbreeding levels is important to avoid problems related to inbreeding depression. This work aims to begin studying the complexity of that problem through an approach that combines linear programming with branch-and-bound to investigate the average execution time for obtaining the problem solution. The results indicate that the problem seems to be computationally intractable.*

1. Introdução

Processos de melhoramento animal dependem de duas estratégias: a seleção e o sistema de acasalamento. A seleção define os animais que se tornarão pais da nova geração do rebanho [Eler 2017]. A escolha pode ser empírica ou usando um índice de seleção que favoreça características desejáveis e desfavoreça problemas na próxima geração [Simões et al. 2020]. O sistema de acasalamento consiste na estratégia de escolha dos pares, para maximizar o valor esperado da futura prole. Sistemas de acasalamento podem visar a produção de animais excepcionais (acasalamento entre iguais) ou uma maior homogeneidade das características do rebanho (acasalamento compensatório).

[Ferreira et al. 2021] apresenta um algoritmo polinomial ótimo para solucionar o problema da maximização do índice esperado da prole, posteriormente usado para mostrar que o problema do acasalamento compensatório também tem solução polinomial,

mantendo a otimalidade da solução [Ferreira 2021]. Em ambos os algoritmos a consanguinidade aparece de forma indireta, por meio da restrição do acasalamento de animais aparentados e limitação do número de vezes que um macho pode acasalar.

O valor máximo da prole, dado um grupo de matrizes e reprodutores, consistiria em cruzar todas as fêmeas com o melhor dos machos, conforme demonstrado matematicamente em [Ferreira et al. 2021], fazendo com que todos os animais da próxima geração sejam irmãos. Cruzamentos consanguíneos são usados para fixação de fenótipos em animais, mas devem ser evitados, visto que o aumento da endogamia causa redução no desempenho de características quantitativas, incluindo reprodução e sobrevivência [Mi et al. 1965].

Embora a existência de relações de parentesco entre animais possa ser usada como restrição, sendo eliminada ou penalizada pelos algoritmos de escolha de acasalamentos, o ideal é que o esquema de cruzamentos não envolva somente o valor do índice, mas também a possibilidade de novos acasalamentos a longo prazo. Ou seja, deve ser possível fazer uma escolha de acasalamentos de maneira que as opções de seleção e cruzamentos sejam mantidas, sem necessidade de compra de novos animais ou sêmen como forma de aumentar a variabilidade genética.

O objetivo deste trabalho é investigar o problema da minimização dos valores de consanguinidade de um rebanho para uma estratégia de acasalamento. A abordagem usa uma estratégia empírica baseada em programação linear [Hillier and Lieberman 2013] combinada com o algoritmo *branch-and-bound* [Sanders et al. 2019]. No restante do texto, apresentaremos os trabalhos relacionados (Seção 2), o detalhamento do método (Seção 3), os resultados alcançados (Seção 4) e a conclusão (Seção 5).

2. Trabalhos relacionados

O método usado para a descoberta das fontes foi a revisão de escopo da literatura [Arksey and O'Malley 2005], com trabalhos a partir de 2015, adicionando-se as referências encontradas na leitura dos trabalhos selecionados.

As abordagens encontradas seguem a mesma estrutura: definição de uma heurística para organização dos acasalamentos seguido de um processo de simulação para análise dos resultados. A exceção (parcial) é [Nomura et al. 2002], que usa programação linear para resolver parte do problema. As heurísticas mais comuns envolvem a definição de grupos de animais e circulação dos machos por esses grupos. As estratégias podem ser circulares, cíclicas, hierárquicas ou usando métodos empíricos clássicos da literatura.

A maior parte dos trabalhos encontrados busca comparar métodos diferentes e já existentes: [Nomura and Yonezawa 1996] compara três métodos heurísticos de organizações de grupos de acasalamento: um modelo circular e dois modelos cíclicos (métodos *Falconer*, *HAN-rotational* e *Cockerham*), [Paiva 2010] simula esquemas de acasalamento usando os métodos empíricos de acasalamento com exclusão de irmãos e meio-irmãos, acasalamento de melhores com melhores, acasalamento de melhores com piores e acasalamento ao acaso, [Pryce et al. 2012] realiza a comparação de diferentes métodos para redução das taxas de endogamia, todas baseadas em informações genéticas dos animais, [Liu et al. 2017] testam a hipótese de que estratégias de acasalamento utilizando dados genotípicos possuem um melhor desempenho na redução das

taxas de endogamia, quando comparados a métodos que utilizam somente informações do pedigree, [Henryon et al. 2009] também compara os sistemas de acasalamento com coancestralidade mínima e minimização da covariância entre as contribuições genéticas ancestrais para seleções por truncamento, [Fernández et al. 2003] compara dois modelos hierárquicos individuais e um geracional, buscando obter a mínima coancestralidade global.

Propostas de novos métodos, ou novas combinações de métodos existentes aparecem em [Honda et al. 2005], que usa uma estrutura hierárquica de raças, com passagem de vantagens genéticas do nível superior para o inferior. [Nomura 1999] combina o método de acasalamento compensatório e modelo de coancestralidade mínima, em que os pares são formados seguindo os conceitos do acasalamento compensatório seguido pela troca das fêmeas nos pares, seguindo os conceitos da mínima coancestralidade, sem que o modelo compensatório seja prejudicado. [Nomura et al. 2002] combina estratégias de melhoria de índices de seleção com taxa de consanguinidade e apresentam uma comparação entre diferentes combinações de métodos de seleção e acasalamento para linhas fechadas de frango.

Os resultados dos trabalhos foram geralmente analisados à luz dos resultados estatísticos dos processos de simulação. Achados mais relevantes indicam que heurísticas de movimento de animais entre grupos levam a resultados bastante similares; que escolhas baseadas em dados genômicos levam a resultados melhores do que escolhas baseadas em pedigree, embora não sejam aplicáveis em larga escala; que modelos que levam em conta explicitamente a coancestralidade dos animais têm resultados melhores; e que acasalamentos compensatórios levam a menores índices de consanguinidade.

Nenhum dos trabalhos encontrados traz uma descrição formal do problema ou busca analisar sua complexidade. Todas as técnicas usadas são heurísticas e baseadas em simulações, sendo que em alguns trabalhos sugere-se o uso de meta-heurísticas. A complexidade computacional dos métodos ou do problema também não é discutida em nenhum trabalho. Nossa abordagem parece ser original em relação à literatura sobre o tema.

3. Material e métodos

Modelos que podem ser descritos por um conjunto de variáveis combinadas por equações e inequações lineares podem ser resolvidos pelo método Simplex [Dantzig 1963], que opera por meio do caminhamento no espaço de soluções viáveis do problema. Quando parte das variáveis somente admite soluções inteiras, usa-se a técnica de relaxamento das restrições (passando a admitir também valores reais para as soluções) seguido pela descoberta das soluções inteiras por métodos como *branch-and-bound* [Lawler 1966], que opera dentro do espaço de soluções viáveis.

O tempo de execução do algoritmo Simplex depende do modelo que está sendo resolvido: se o problema for computacionalmente intratável, o tempo de execução é exponencial no número de variáveis do problema [Lenstra and Kan 1983]. A medição empírica do tempo de execução da combinação Simplex e *branch-and-bound* pode ser um indicativo da complexidade do problema que está sendo resolvido. Essa estratégia foi usada por [Motta 2021], e apresentou a correta indicação de que o problema da

maximização do índice (valor) da prole poderia ser resolvido em tempo polinomial¹.

Problemas de programação linear pertencem à classe de complexidade NP, que contém os problemas que podem ter uma possível solução verificada em tempo polinomial e a restrição do problema para soluções inteiras é NP-completa [Garey and Johnson 1979]. Como o complemento do problema da programação linear também está em NP, ele somente será NP-completo se $P=NP^2$ [Arora and Barak 2009].

Este trabalho faz uso de um procedimento experimental de análise da complexidade média da execução do modelo linear descrito na Definição 1, com os algoritmos Simplex e *branch-and-bound*, implementados no pacote R *lpSolve* [Berkelaar et al. 2020]. O algoritmo implementado evolui o algoritmo BrangusSeletion [Motta et al. 2021] para incorporar a consanguinidade ao modelo, além do valor da futura prole dado pelo índice de seleção. O pacote R *pedigreemm* [Bates and Vazquez 2015] é usado para o cálculo dos valores de coancestralidade e de consanguinidade do rebanho, a partir dos dados de parentesco dos animais.

Definição 1 (Modelo linear do sistema de acasalamento) *Seja M o conjunto de reprodutores e F o conjunto de matrizes selecionadas para acasalamento. O modelo do sistema de acasalamento é expresso pela função objetivo*

$$MIN y = \sum_{i \in M} \sum_{j \in F} v_{ij} x_{ij} \quad (1)$$

em que cada variável x_{ij} pode assumir um valor 1 ou 0, consoante o par ij deva ou não acasalar (ou seja, pertença ou não à solução do problema) e o valor v_{ij} é a contribuição ponderada de melhoramento calculada como

$$v_{ij} = \frac{\lambda(m_i, f_i) \cdot P_c - \pi(m_i, f_i) \cdot P_e}{P_c + P_e} \quad (2)$$

em que $\lambda(m_i, f_i)$ e $\pi(m_i, f_j)$ são, respectivamente, o índice de consanguinidade e o valor do índice de seleção da prole resultante do par $(m_i, f_j) \in M \times F$. Os valores $P_e, P_c \in \mathbb{R}$ são os pesos relativos ao índice de seleção e ao índice de coancestralidade, respectivamente.

As restrições resultam diretamente da definição do problema:

$$\forall j \in F \sum_{i \in M} x_{ij} = 1 \quad (3)$$

$$\forall i \in M \sum_{j \in F} x_{ij} \leq t_{\max}(i) \quad (4)$$

$$\forall i \in M \sum_{j \in F} x_{ij} \geq t_{\min}(i) \quad (5)$$

¹Um algoritmo é dito eficiente quando seu tempo de execução pode ser limitado superiormente por uma função polinomial no tamanho da sua entrada [Garey and Johnson 1979].

²A questão $P = NP$ é considerada o maior problema em aberto da Computação e pode ser resumida como “Todo problema que pode ter sua solução verificada em tempo polinomial também pode ser resolvido em tempo polinomial?”. O atual conhecimento indica que não, o que implicaria que problemas NP-completos somente admitem algoritmos exponenciais (i.e., são computacionalmente intratáveis).

onde $t_{\max}(i)$ e $t_{\min}(i)$ são os limites mínimo e máximo do uso de cada macho na atribuição de pares.

Para um conjunto de n matrizes e m reprodutores, o modelo linear acima possui $n \cdot m$ variáveis e $n + 2m$ restrições. O cálculo do valor de coancestralidade λ entre os animais é executado uma única vez e acessado em tempo constante quando necessário. Os coeficientes π são calculados como a média entre os valores do índices dos pais. Os valores P_c e P_e indicam, respectivamente, a importância relativa entre a minimização da coancestralidade e a maximização do valor do índice usado para valoração da prole, visto que o aumento de um quase sempre significa a diminuição do outro [Weigeland and Lin 2002].

A base de dados usada foi fornecida pela Embrapa Pecuária Sul e tem 3.515 bovinos da raça Brangus (1.468 machos e 2.029 fêmeas) com índice de seleção já calculado. Os testes foram realizados com escolha aleatória dos animais da base. A estimativa da complexidade média da execução do algoritmo foi feita com 10 execuções para cada configuração de tamanho de entrada e proporção macho/fêmea, considerando diferentes valores de ponderação. O ajuste do número mínimo e máximo de machos usados na solução do problema foi feito de modo a garantir a existência de pelo menos uma solução.

4. Discussão dos resultados

A estimativa da complexidade média da execução dos algoritmos Simplex e *branch-and-bound* foi obtida pela execução sistemática dos algoritmos, com diferentes proporções entre machos e fêmeas (nomeadamente 1:2, 1:5, 1:10 e 1:18) e com incrementos da quantidade de animais envolvidos para que uma curva do tempo de execução pudesse ser traçada. Para a configuração 1:2, o número de machos variou de 10 a 100 e o número de fêmeas variou de 20 a 200. Para a configuração 1:5, o número de machos variou de 5 a 55 e o número de fêmeas variou de 25 a 275. Na configuração 1:10, o número de machos variou de 20 a 55 e o número de fêmeas variou de 200 a 550. Por fim, para a configuração 1:18, o número de machos variou de 5 a 40 e o número de fêmeas variou de 90 a 720.

As diferentes ponderações para os valores relativos consanguinidade/valor foram feitas para buscarmos entender como um dos problemas (sabidamente polinomial) se comportava em relação ao outro, cuja complexidade queremos investigar. Para cada proporção machos/fêmeas foram utilizadas seis configurações diferentes de ponderação (P_e e P_c na Eq. 2), variando entre 0 e 1, com incrementos de 0.2.

Os resultados das execuções do algoritmo estão representados na Figura 1. Cada uma das imagens mostra as curvas do tempo médio de execução para valores (P_e, P_c) de $(0, 1)$, $(0.2, 0.8)$, $(0.4, 0.6)$, $(0.6, 0.4)$, $(0.8, 0.2)$ e $(1, 0)$. No primeiro caso somente é dada importância ao valor do índice da prole e no último somente à consanguinidade. Em todas as proporções entre machos e fêmeas usadas, o comportamento do tempo de execução é bastante similar: quanto maior a proporção de importância da consanguinidade, maior é o tempo de execução do algoritmo. A interrupção das linhas significa que o tempo de execução para as instâncias ultrapassou 20 minutos, tendo sido a execução interrompida e o valor não computado.

O formato da curva do tempo de execução nos casos em que a consanguinidade é o fator determinante indica que o algoritmo não é capaz de lidar com entradas de tamanho muito grande, tendo o formato de uma função exponencial. A maior suavidade das curvas

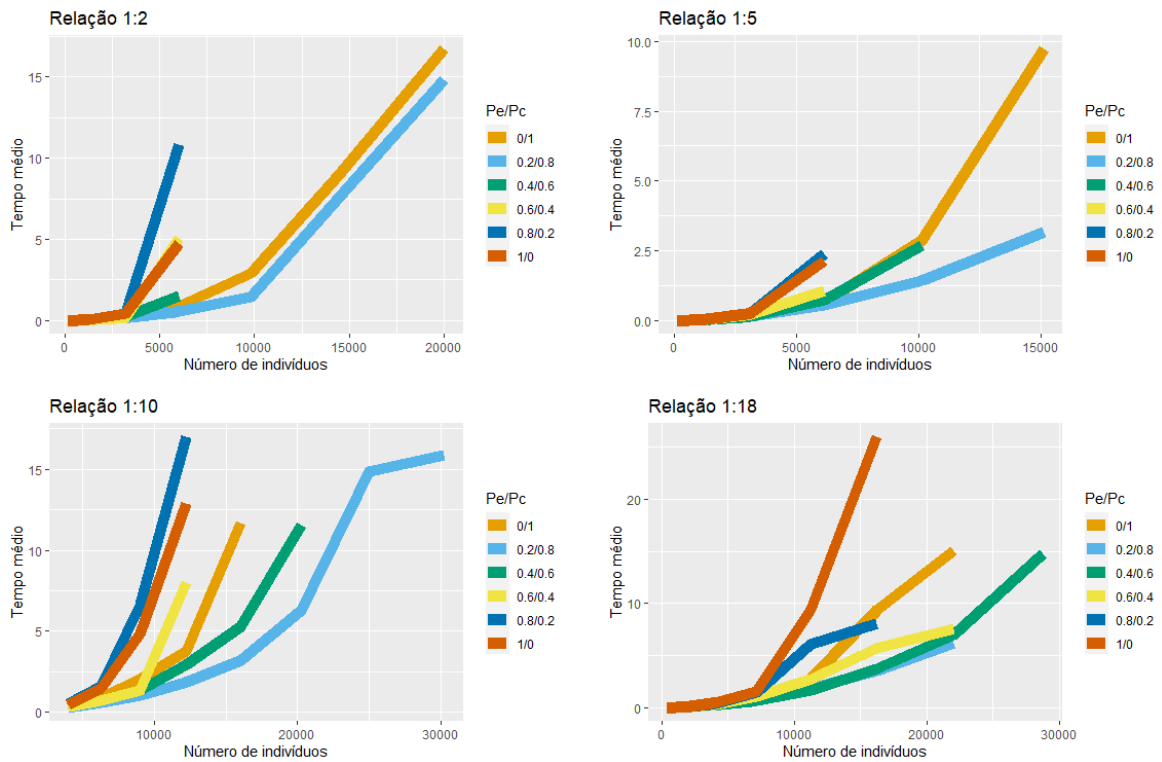


Figura 1. Tempo médio de execução do algoritmo

em que o valor do índice é o elemento preponderante evidencia que o tempo médio, nesse caso, é regido por um polinômio.

A intratabilidade do problema da minimização da consanguinidade não pode ser provada dessa maneira, mas o indicativo é importante e compatível com as características do problema. Intuitivamente, não é possível descobrir a prole que minimiza a consanguinidade sem investigar todas as proles possíveis. Como uma prole pode ser formalizada por uma função $p : F \rightarrow M$ (em que é atribuído um macho para cada fêmea) e existem $|M|^{|F|}$ funções possíveis entre os conjuntos F e M , não é computacionalmente viável investigar todas as proles. Diferentemente do caso do valor esperado da prole, em que uma troca de casais não altera o valor total da prole [Ferreira et al. 2021], a eventual alteração dos pais faz com que a coancestralidade entre a nova prole e rebanho e entre a nova prole com ela mesma precise ser recalculada. Essa diferença parece ser determinante na complexidade dos dois problemas.

5. Conclusão

A minimização da consanguinidade de um rebanho é essencial para garantir a variabilidade genética e evitar a depressão endogâmica. Sistemas de acasalamento precisam levar em conta tanto os ganhos genéticos e fenotípicos do rebanho como também controlar a endogamia. As abordagens na literatura usam heurísticas para determinar os pares que devem acasalar, mas essa abordagem somente é necessária quando o problema é reconhecidamente intratável.

Este trabalho analisou a complexidade dos métodos Simplex e *branch-and-bound* para um modelo de programação inteira que visa determinar como os pares selecionados

devem acasalar. Fez-se uso de ponderação entre o valor econômico e a consanguinidade do rebanho. Os resultados apontam o problema da consanguinidade mínima ser computacionalmente intratável.

Os próximos passos deste trabalho envolvem construir a prova de que o problema pertence à classe de complexidade NP-difícil (problemas de otimização cujo problema de decisão equivalentemente é NP-completo). Adicionalmente, pretendemos construir um algoritmo (que use heurísticas ou meta-heurísticas) que tenha como entrada uma solução ótima do problema da maximização do valor econômico do rebanho. A investigação das propriedades do problema, para viabilizar uma boa heurística, também é um caminho a seguir.

Referências

- Arksey, H. and O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *Int. J. Social Research Methodology*, 8(1):19–32.
- Arora, S. and Barak, B. (2009). *Computational Complexity: a Modern Approach*. Cambridge University Press, Cambridge, UK.
- Bates, D. and Vazquez, A. I. (2015). pedigreeemm: Pedigree-based mixed-effects models. R package version 0.3-3.
- Berkelaar, M. et al. (2020). Interface to lp_solve v. 5.5 to solve linear/integer programs. *R package*.
- Dantzig, G. B. (1963). Linear programming and extensions. *Princeton University Press*.
- Eler, J. P. (2017). *Teorias e métodos em melhoramento genético animal: seleção*. Faculdade de Zootecnia e Engenharia de Alimentos da USP, Pirassununga – São Paulo.
- Fernández, J., Toro, M., and Caballero, A. (2003). Fixed contributions designs vs. minimization of global coancestry to control inbreeding in small populations. *Genetics*, 165(2):885–894.
- Ferreira, A. P. L. (2021). On the problem of compensatory mating in animal breeding. In *Anais do VI Workshop-Escola de Informática Teórica (WEIT 2021)*, pages 96–103, Bagé, RS.
- Ferreira, A. P. L., Yokoo, M. J.-I., and Motta, B. E. T. (2021). On the problem of optimal mating in animal breeding. In *Simposio Latinoamericano de Teoría Computacional, Proceedings of the XLVII Conferencia Latinoamericana de Informática (CLEI 2021)*, pages 1–7, San José, Costa Rica. IEEE.
- Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Series of Books in the Mathematical Sciences. W. H. Freeman, 1st edition.
- Henryon, M., Sørensen, A., and Berg, P. (2009). Mating animals by minimising the covariance between ancestral contributions generates less inbreeding without compromising genetic gain in breeding schemes with truncation selection. *Animal*, 3(10):1339–1346.
- Hillier, F. S. and Lieberman, G. J. (2013). *Introdução à pesquisa operacional*. McGraw-Hill Brasil, Porto Alegre.

- Honda, T., Nomura, T., and Mukai, F. (2005). Prediction of inbreeding in commercial females maintained by rotational mating with partially isolated sire lines. *Journal of Animal Breeding and Genetics*, 122(5):340–348.
- Lawler, E. L. (1966). Branch and bound methods for integer programming. *INFORMS Journal on Computing*, 7(1):73–87.
- Lenstra, A. and Kan, S. (1983). Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8(4):538–548.
- Liu, H., Henryon, M., and Sørensen, A. (2017). Mating strategies with genomic information reduce rates of inbreeding in animal breeding schemes without compromising genetic gain. *Animal*, 11(4):547–555.
- Mi, M. P., Chapman, A. R., and Tyler, W. J. (1965). Effect of mating system on production traits in dairy cattle. *J. Dairy Sci.*, 48:77–84.
- Motta, B. E. T. (2021). BrangusSelection: algoritmo de otimização para seleção de acasalamentos de bovinos com foco na maximização dos ganhos financeiros. Master's thesis, Programa de Pós-graduação em Computação Aplicada, Universidade Federal do Pampa.
- Motta, B. E. T., Ferreira, A. P. L., and Yokoo, M. J.-I. (2021). BrangusSelection: um algoritmo para seleção ótima de acasalamentos com índice de seleção customizável. In *Anais do XIII Congresso Brasileiro de Agroinformática*, pages 35–43, Porto Alegre. Sociedade Brasileira de Computação.
- Nomura, T. (1999). A mating system to reduce inbreeding in selection programmes: theoretical basis and modification of compensatory mating. *Journal of Animal Breeding and Genetics*, 116(5):351–361.
- Nomura, T., Yamaguchi, S., Mukai, F., and Yamamoto, A. (2002). Optimization of selection and mating schemes in closed broiler lines. *Animal science journal*, 73(6):435–443.
- Nomura, T. and Yonezawa, K. (1996). A comparison of four systems of group mating for avoiding inbreeding. *Genetics Selection Evolution*, 28(2):141–159.
- Paiva, A. L. C. (2010). Avaliação de métodos de controle de endogamia utilizando dados simulados.
- Pryce, J., Hayes, B., and Goddard, M. (2012). Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. *Journal of dairy science*, 95(1):377–388.
- Sanders, P., Mehlhorn, K., Dietzfelbinger, M., and Dementiev, R. (2019). *Generic Approaches to Optimization*, pages 357–392. Springer International Publishing, Cham.
- Simões, M. R. S., Leal, J. J. B., Minho, A. P., Gomes, C. C., MacNeil, M. D., Costa, R. F., Junqueira, V. S., Schmidt, P. I., Cardoso, F. F., Boligon, A. A., and Yokoo, M. J. (2020). Breeding objectives of Brangus cattle in Brazil. *Journal of Animal Breeding and Genetics*, 137(2):177–188.
- Weigeland, K. and Lin, S. (2002). Controlling inbreeding by constraining the average relationship between parents of young bulls entering AI progeny test programs. *Journal of dairy science*, 85(9):2376–2383.