

Uma Análise Comparativa de Métricas de Pontuação e Procedimentos de Busca aplicados ao Algoritmo de Otimização Bayesiano

Ciniro Aparecido Leite Nametala

Algoritmos de Estimação de Distribuição

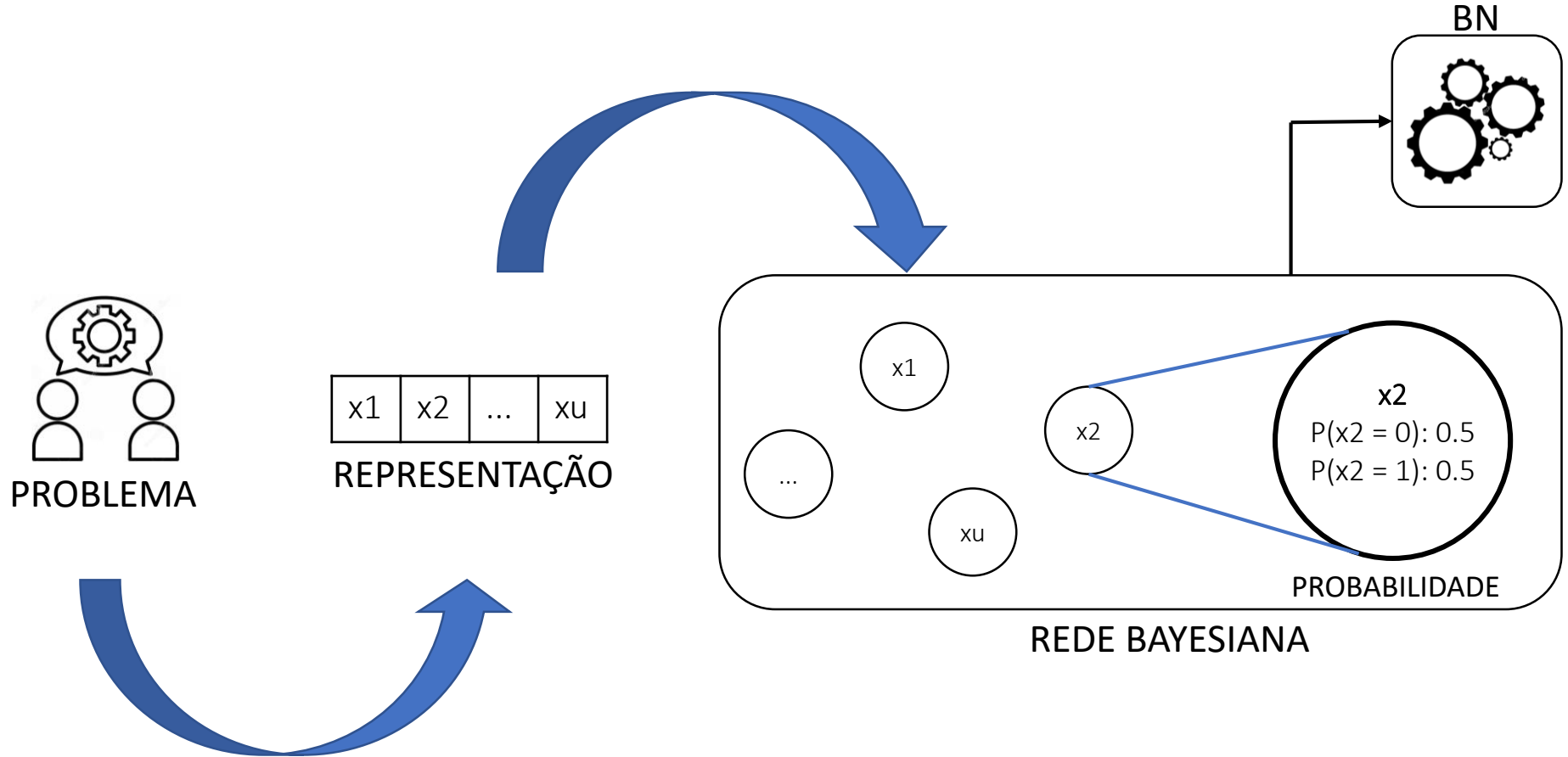
Prof. Dr. Alexandre Cláudio Botazzo Delbem

Programa de Pós-Graduação em Engenharia Elétrica

São Carlos, 29 de Novembro de 2019



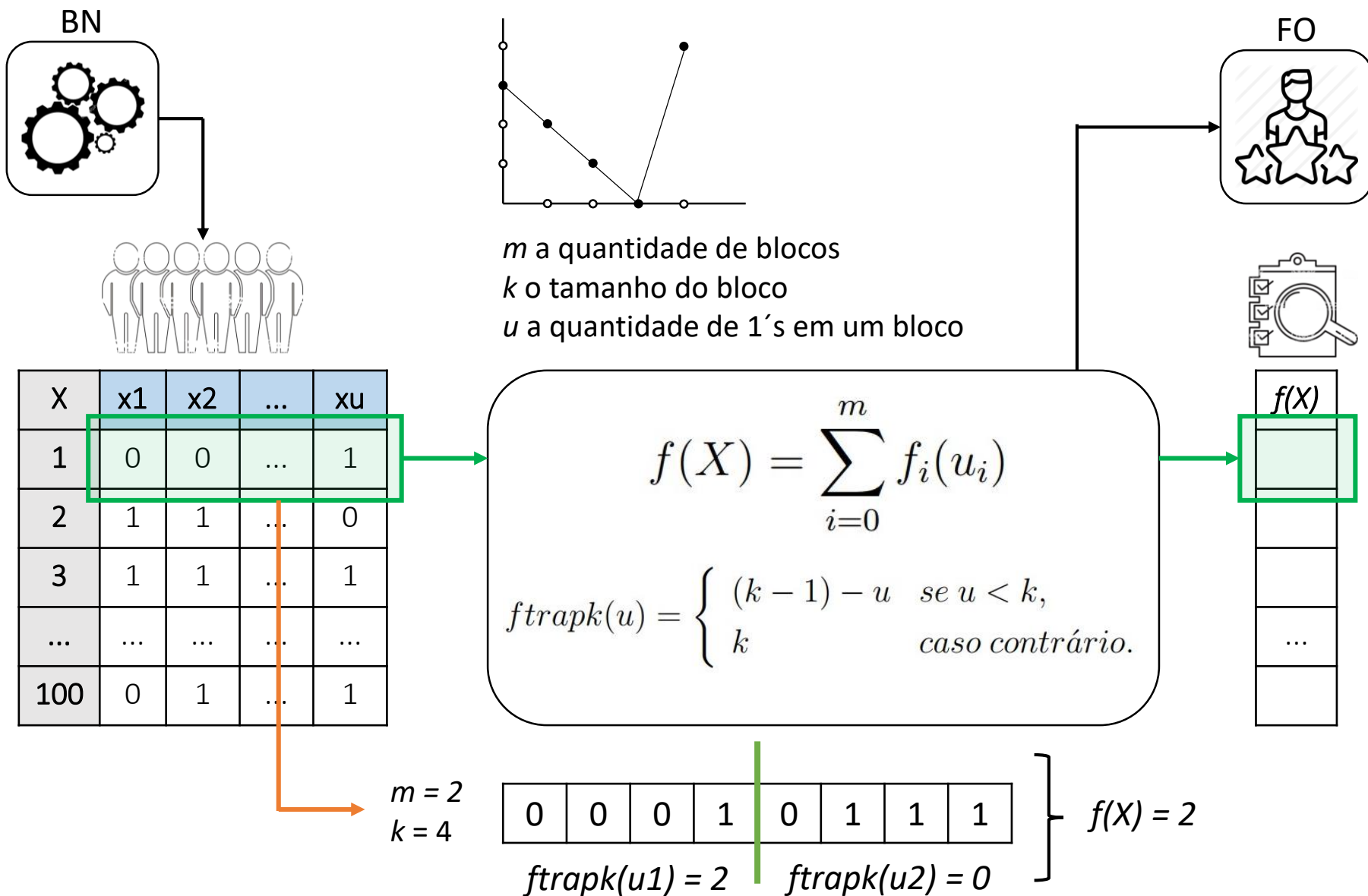
Inicialização: Rede Bayesiana e População Inicial



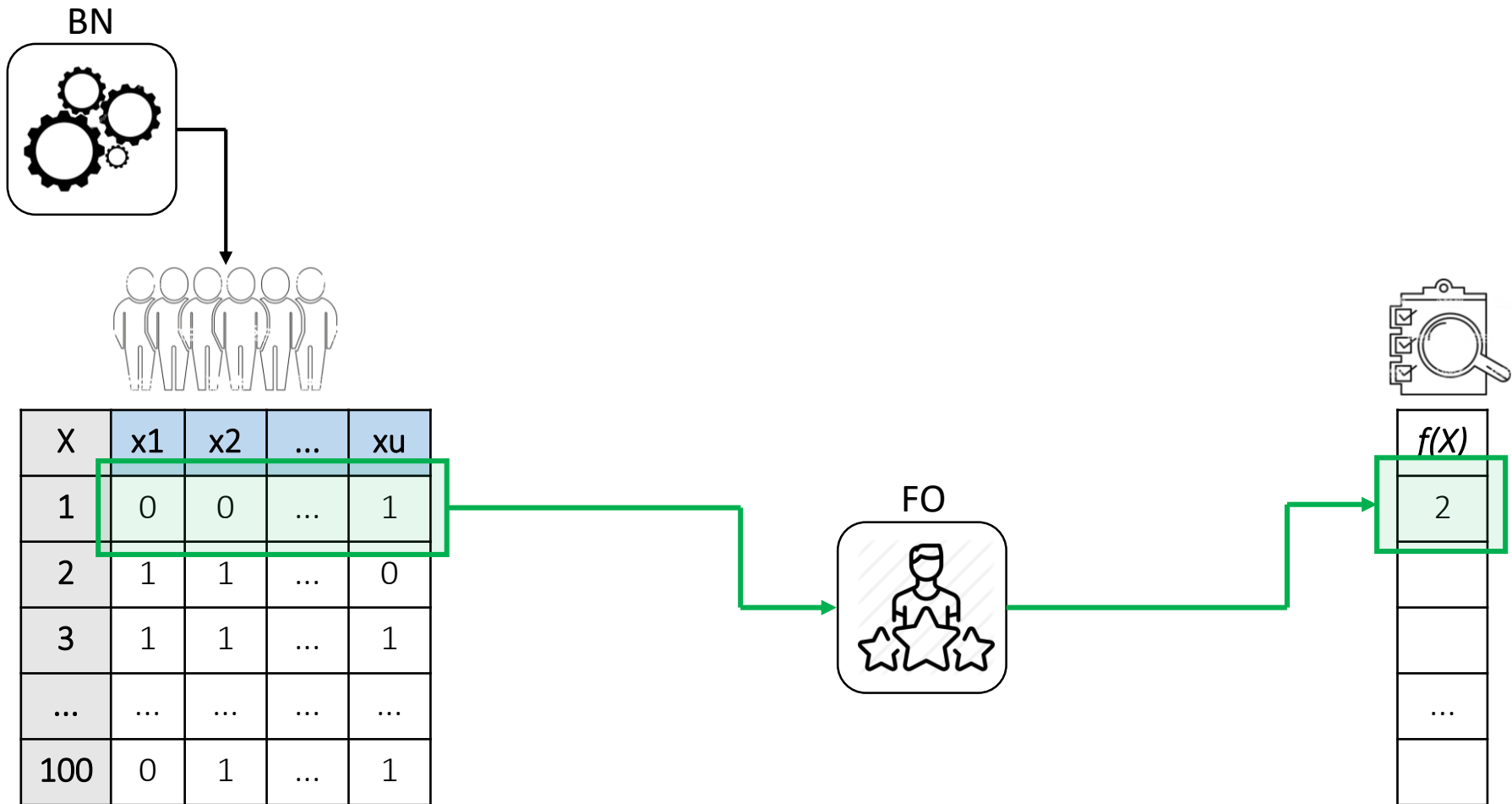
Representação binária

Grafos direcionais acíclicos

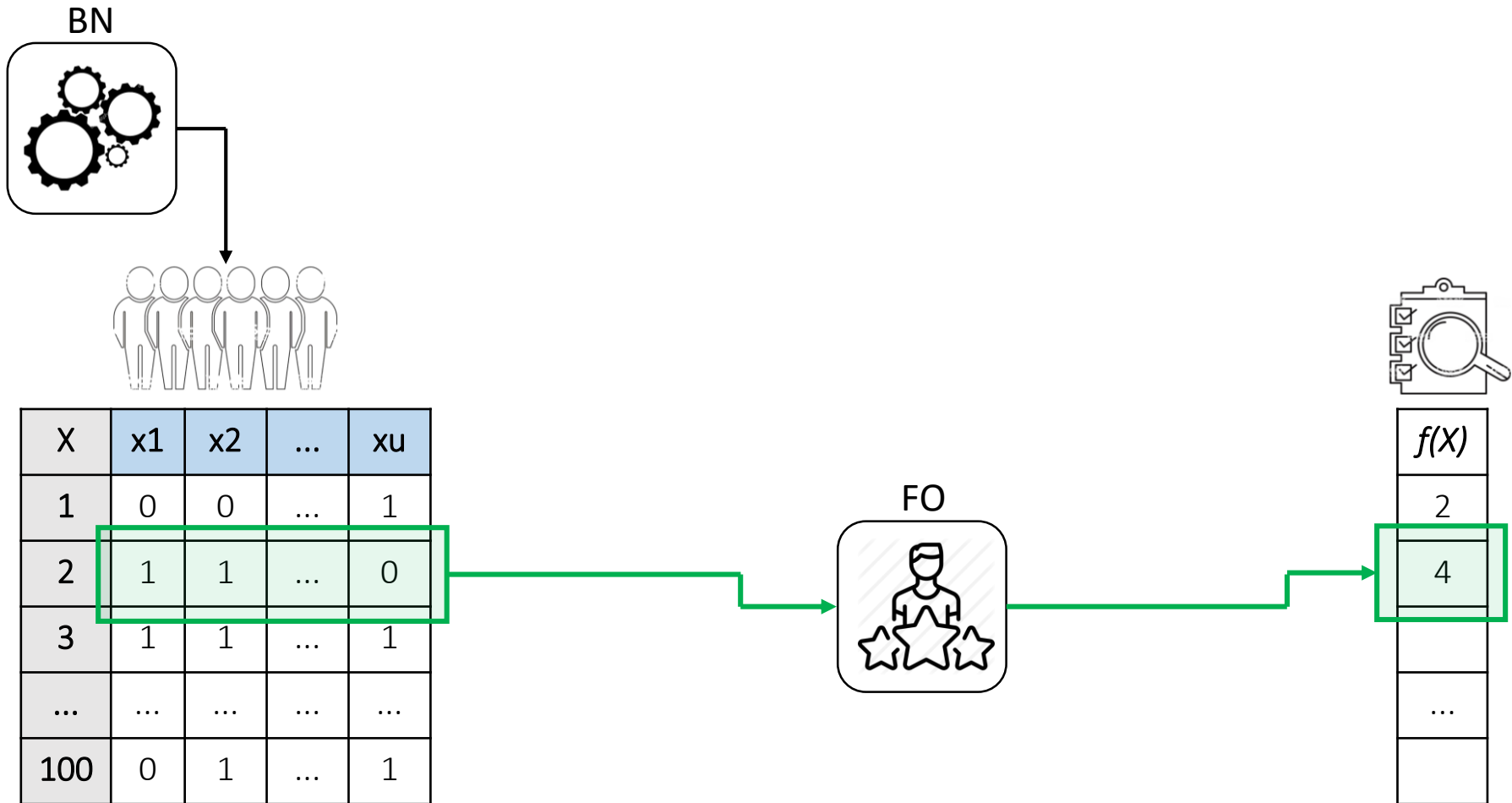
Função Objetivo: TrapK



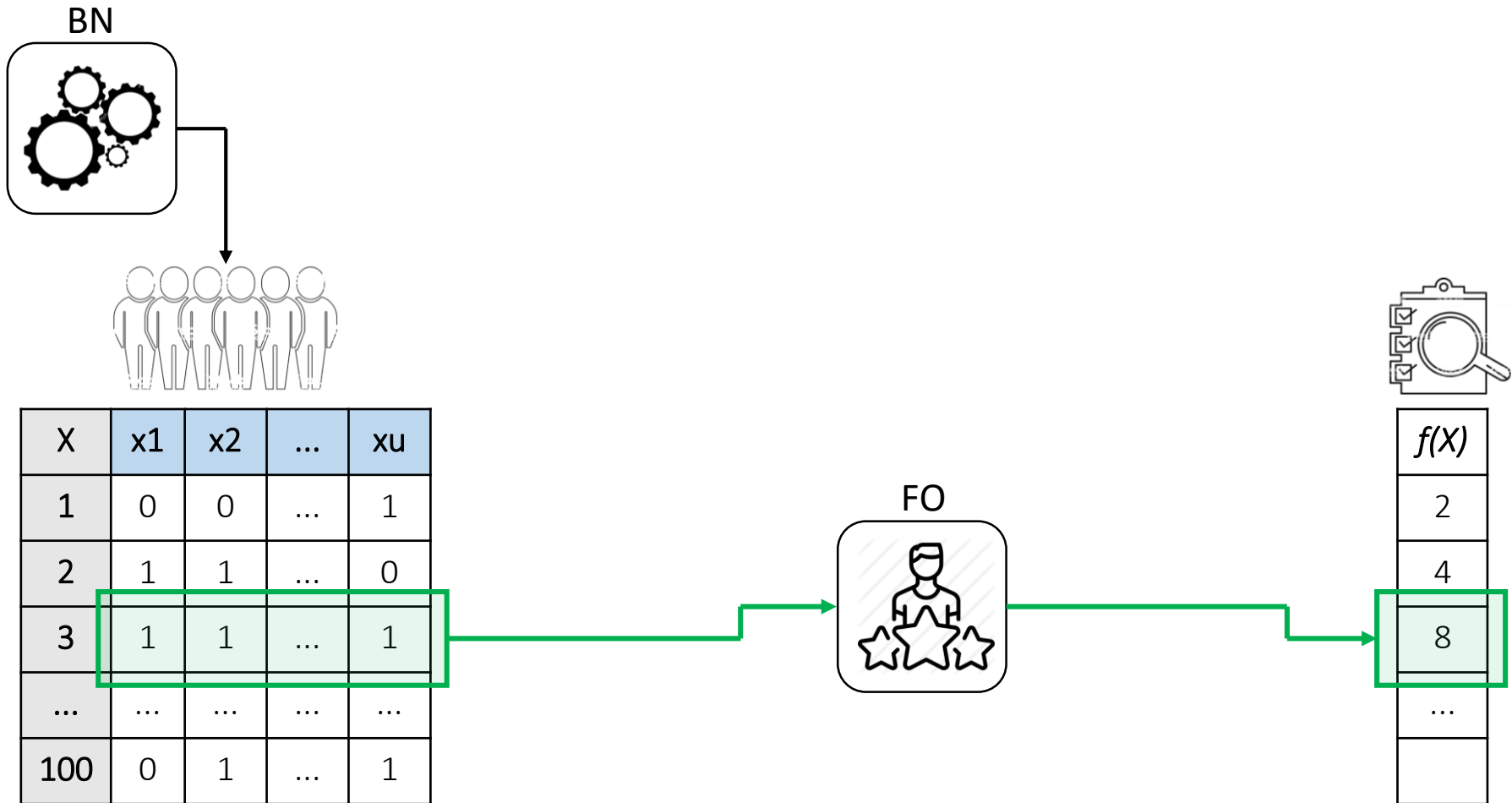
Avaliação da População Inicial



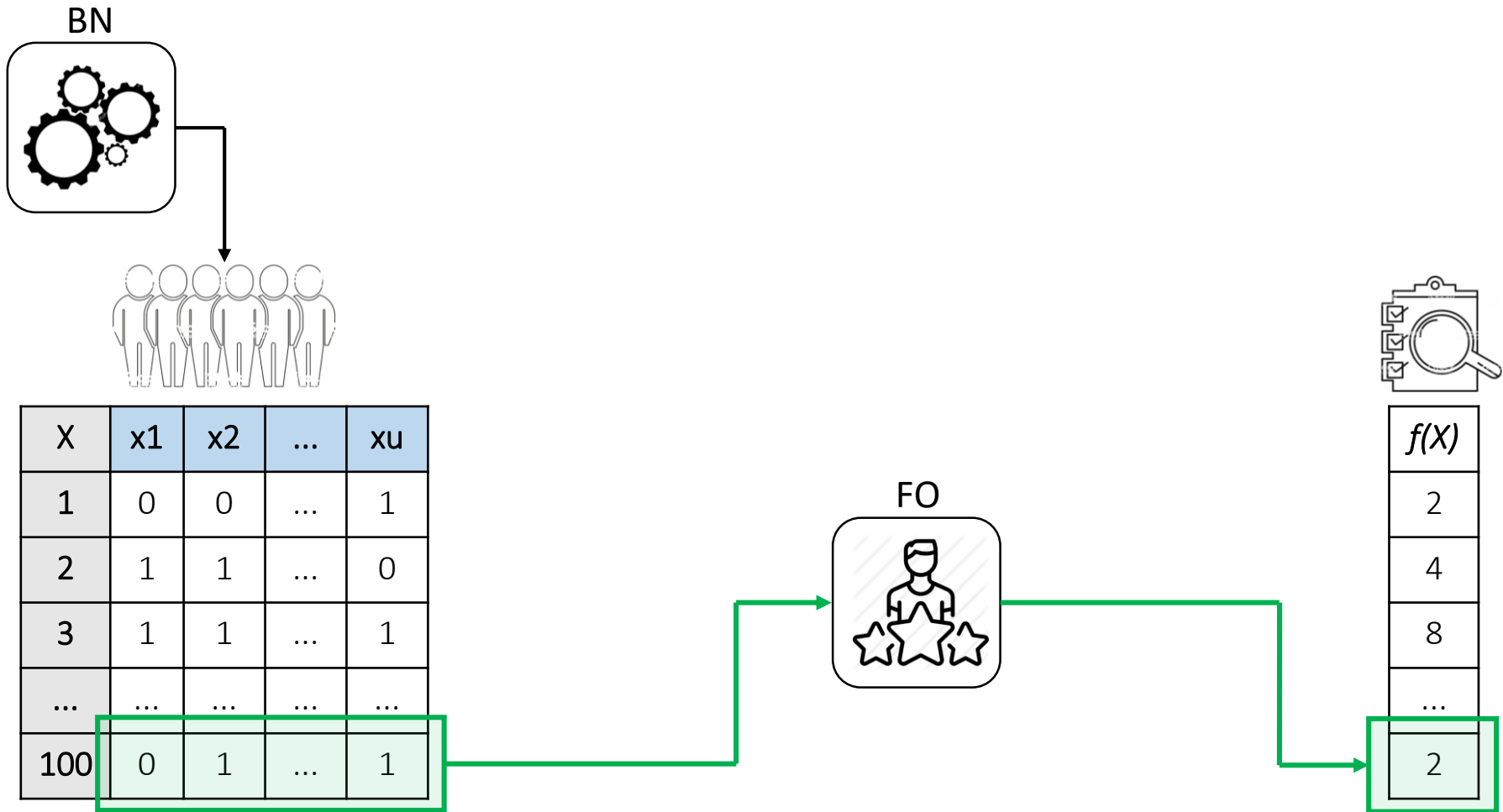
Avaliação da População Inicial



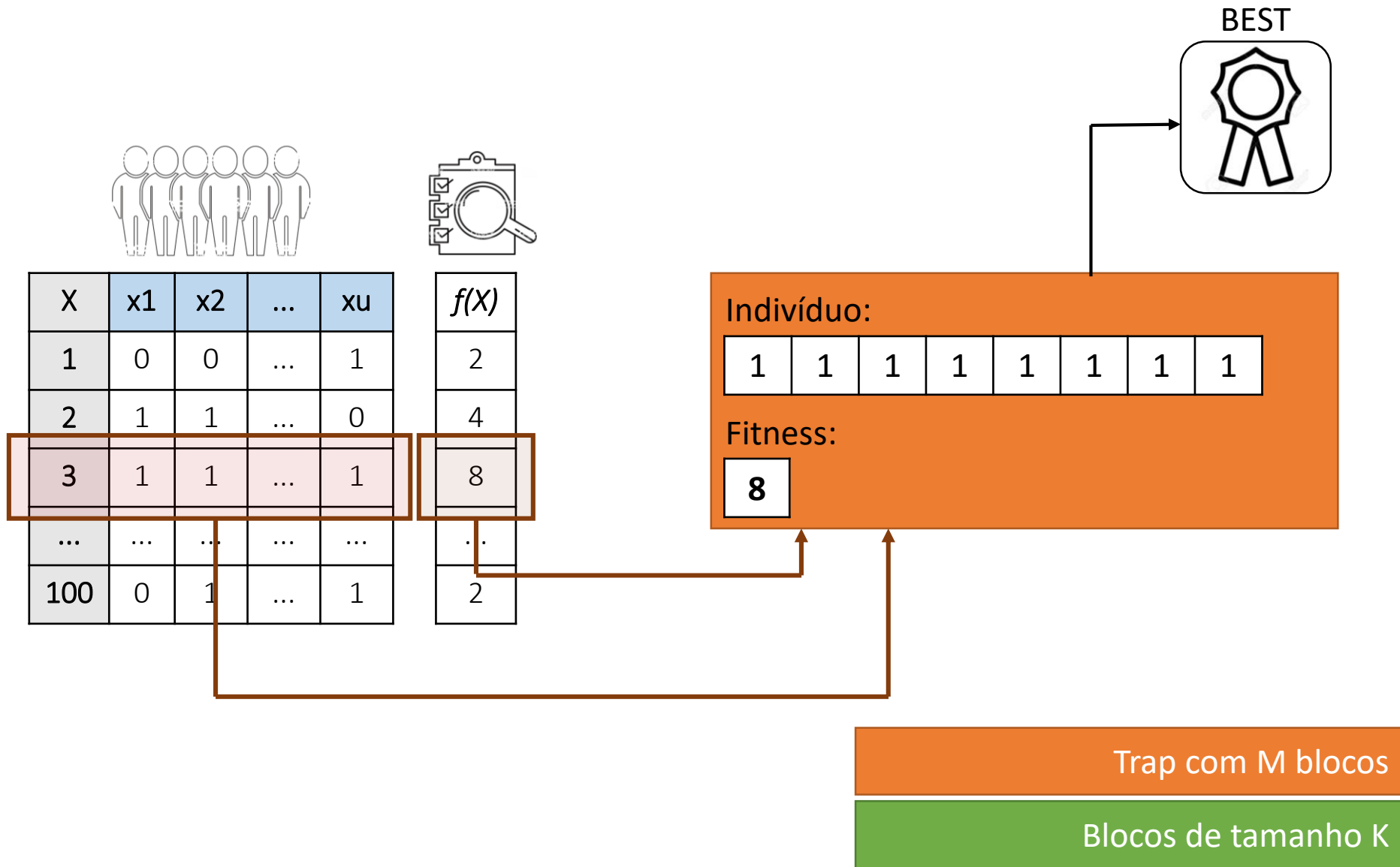
Avaliação da População Inicial



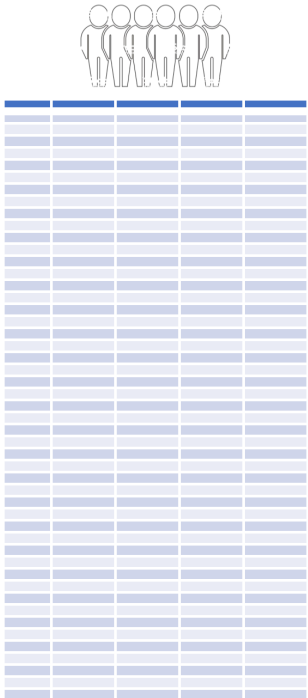
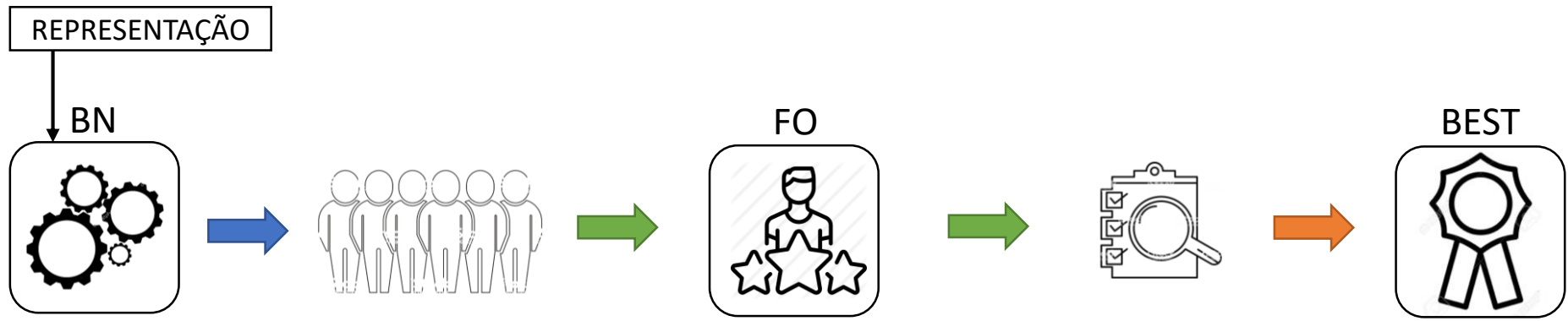
Avaliação da População Inicial



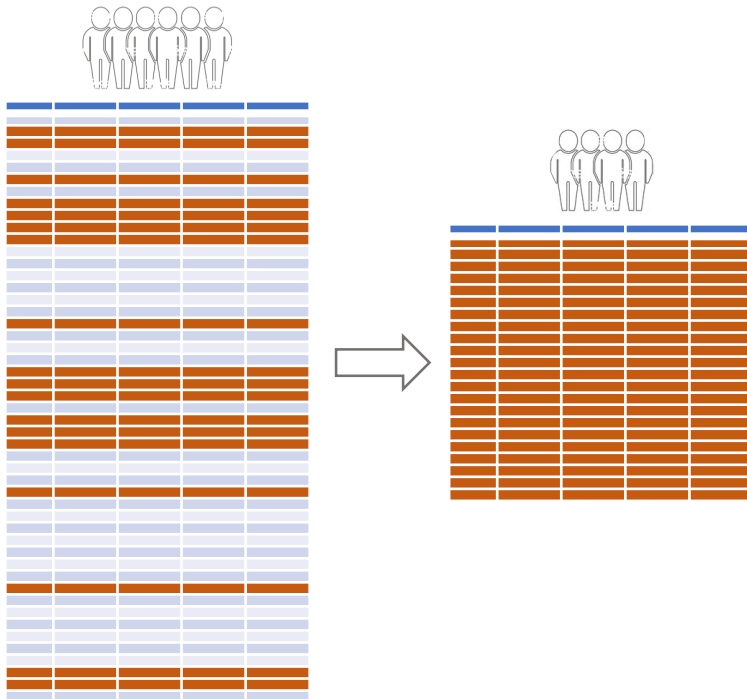
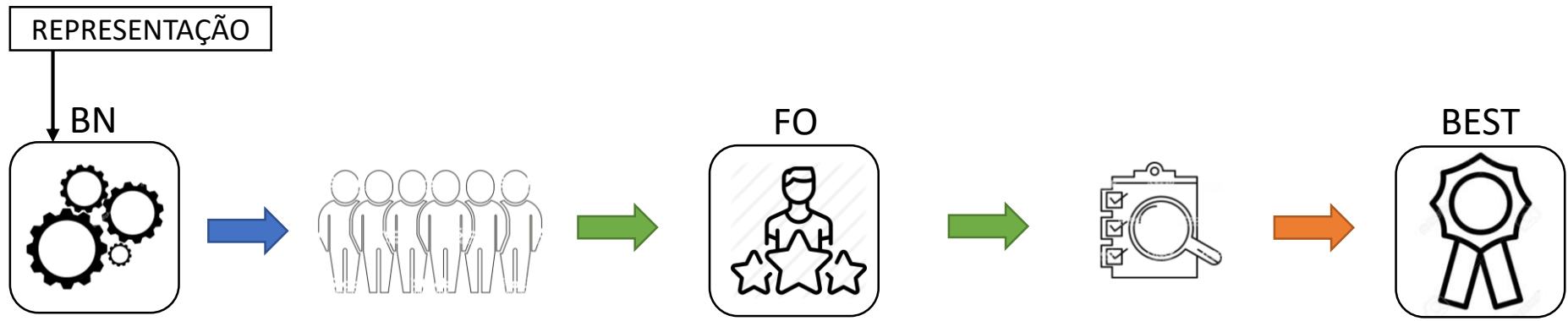
Armazenamento do Melhor



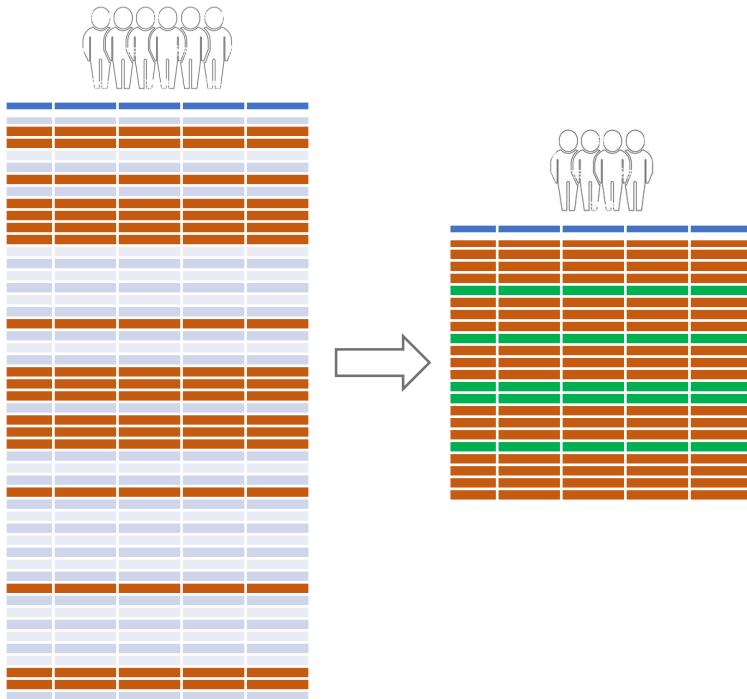
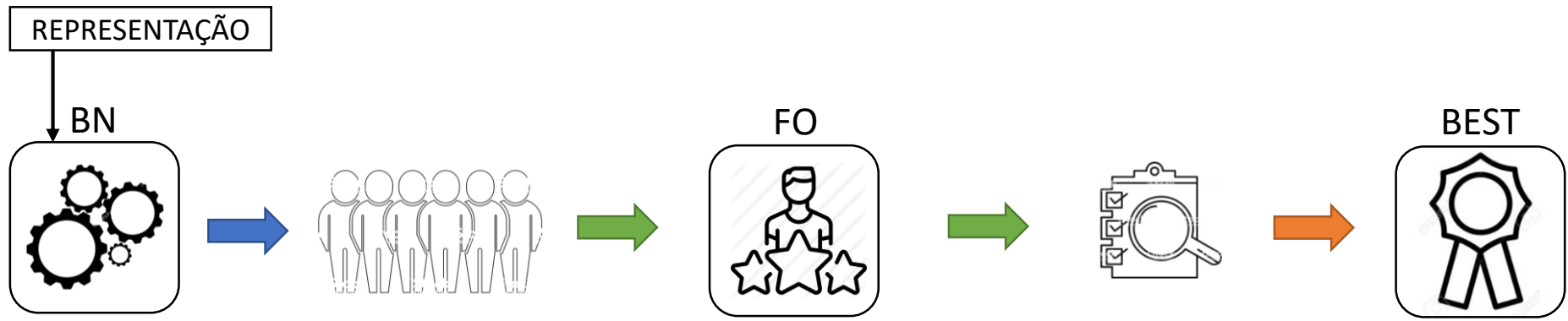
Torneio iterativo



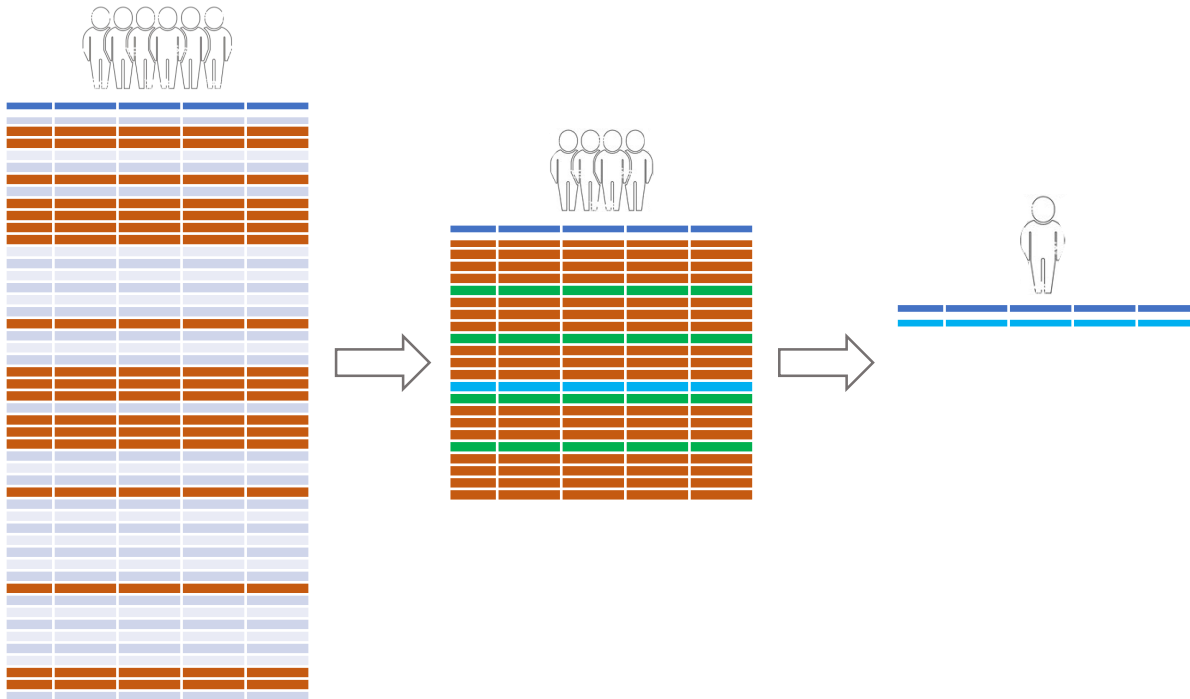
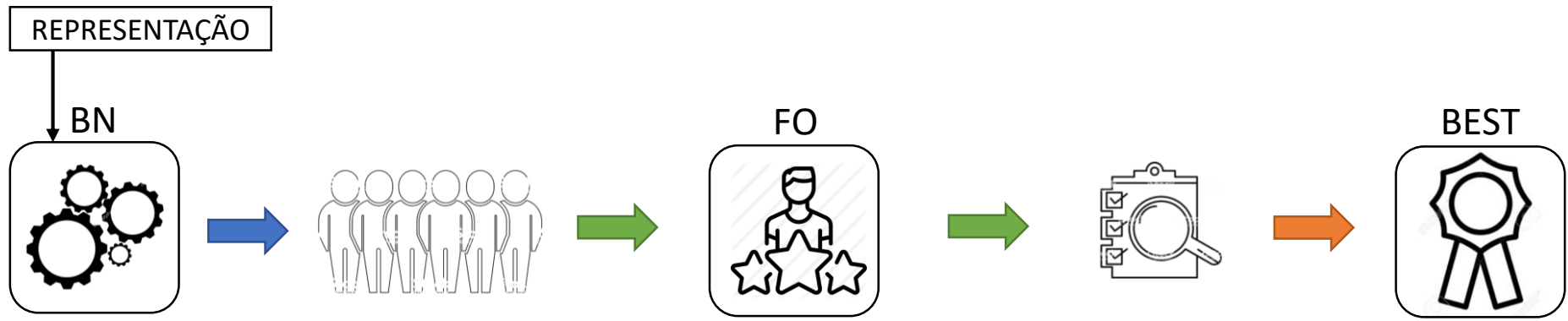
Torneio iterativo



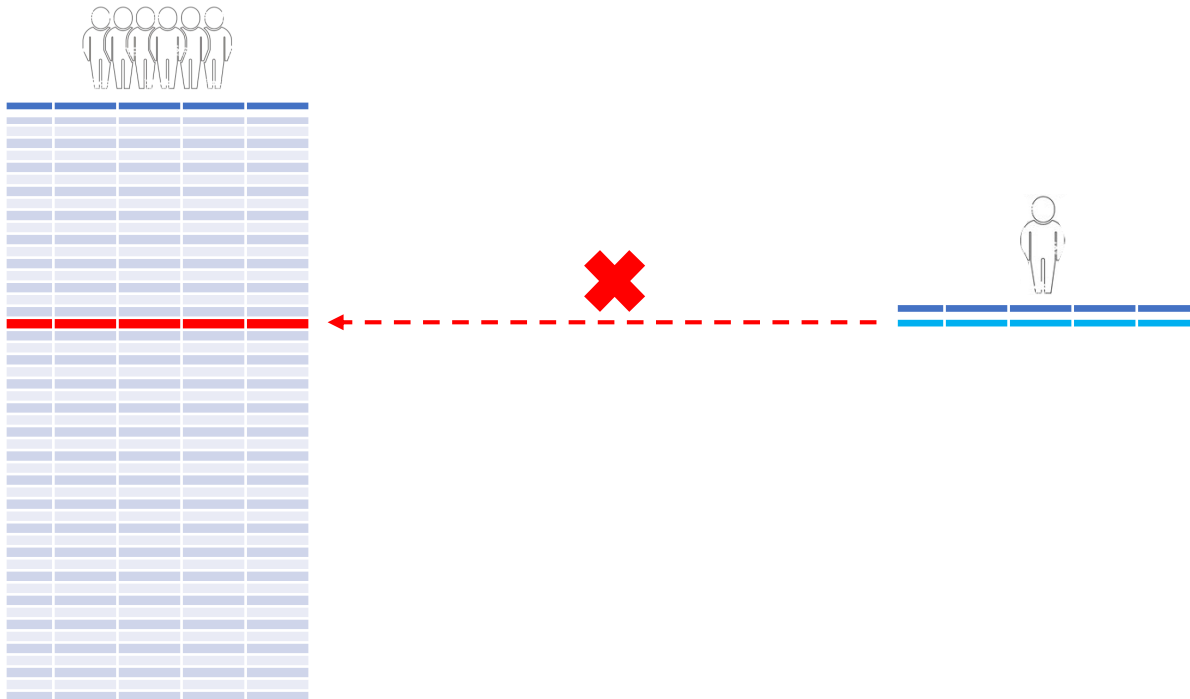
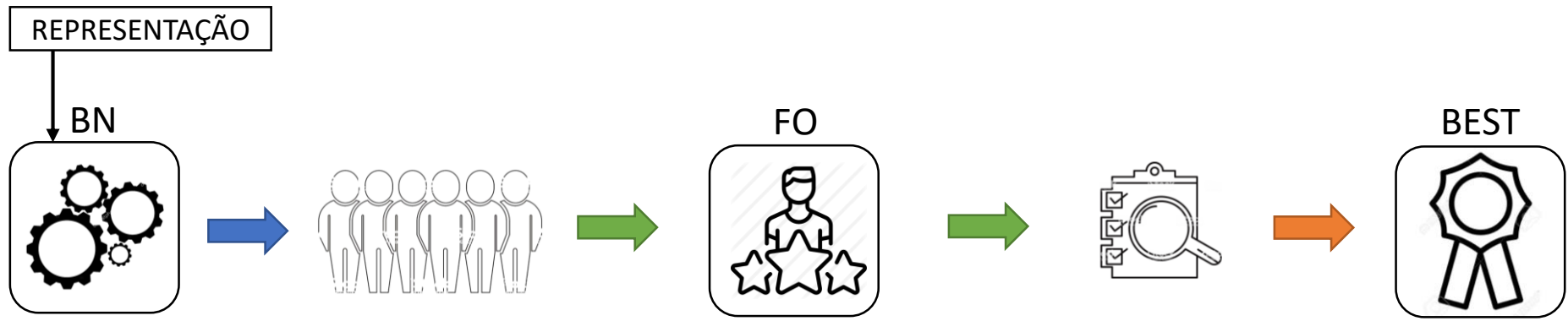
Torneio iterativo



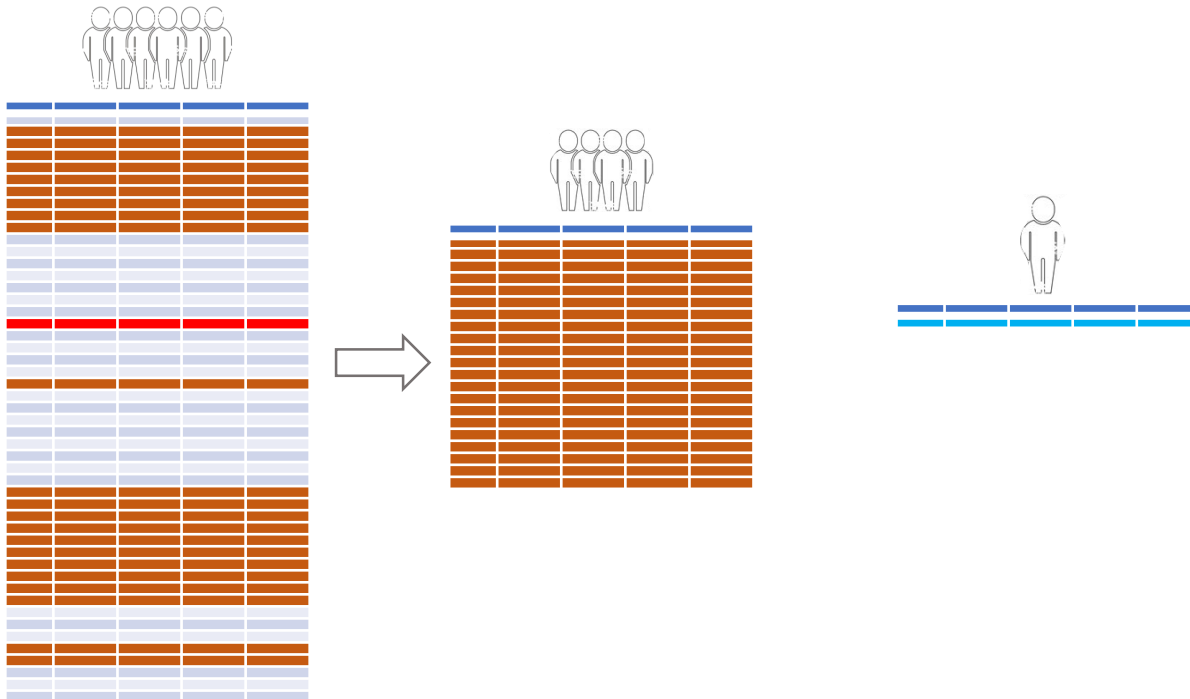
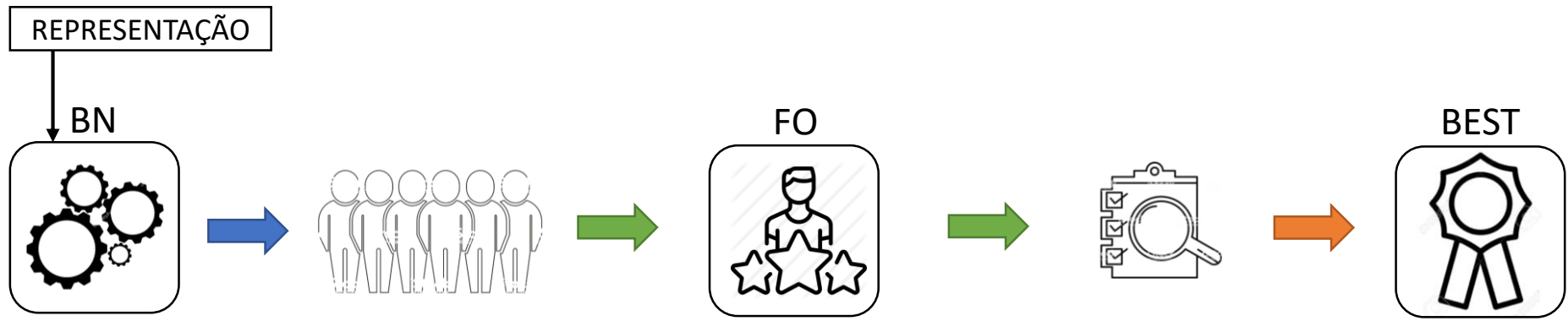
Torneio iterativo



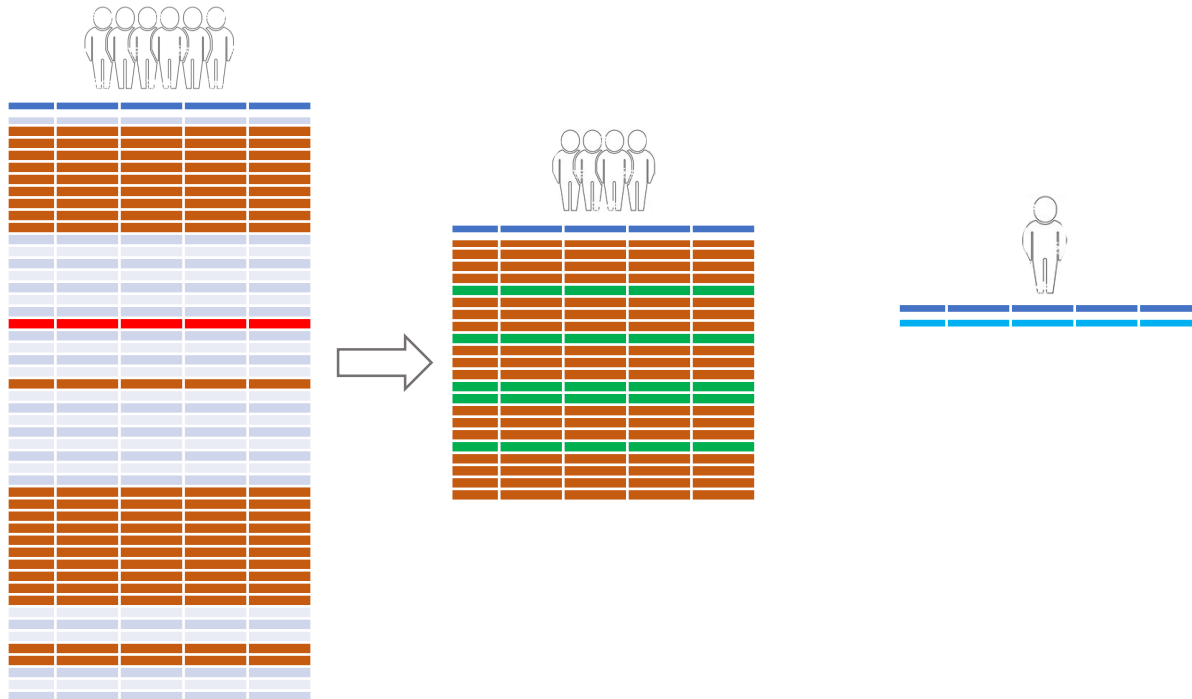
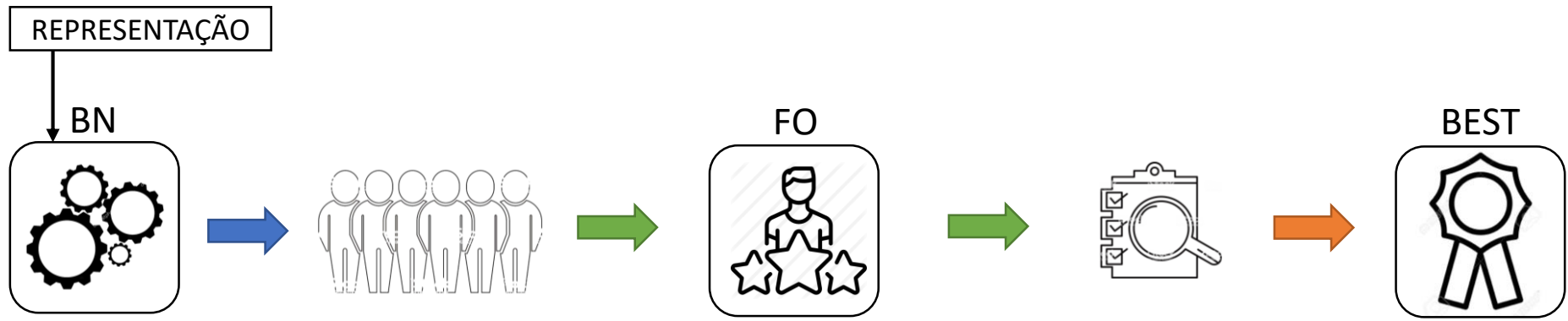
Torneio iterativo



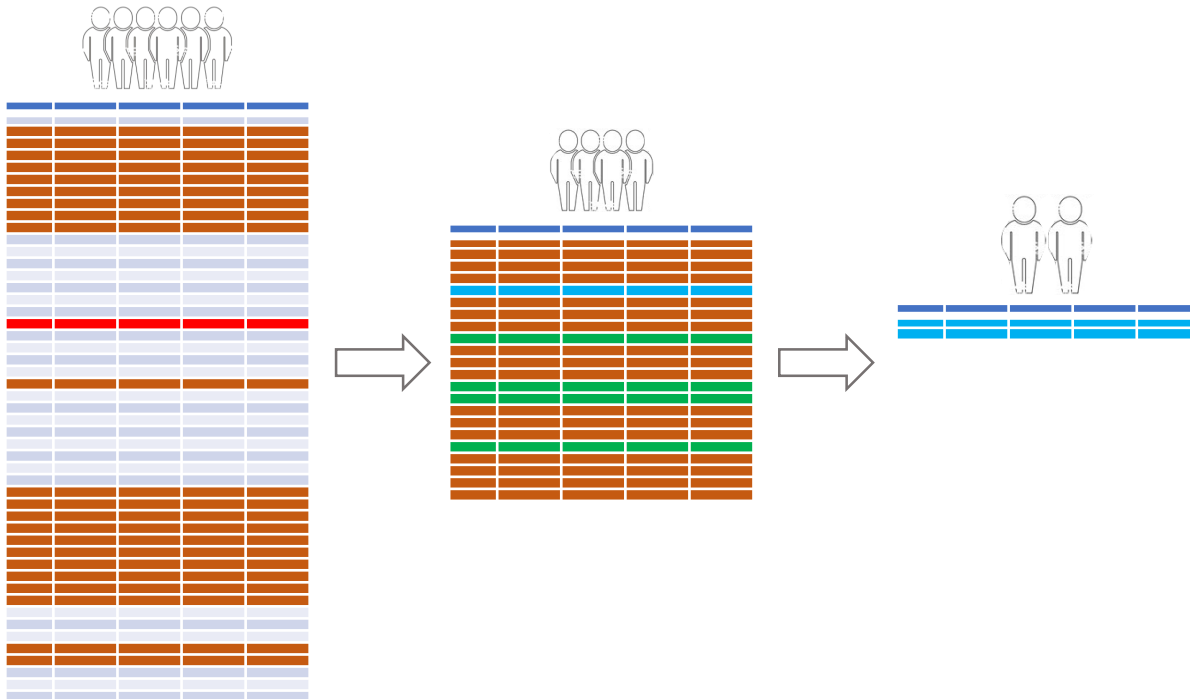
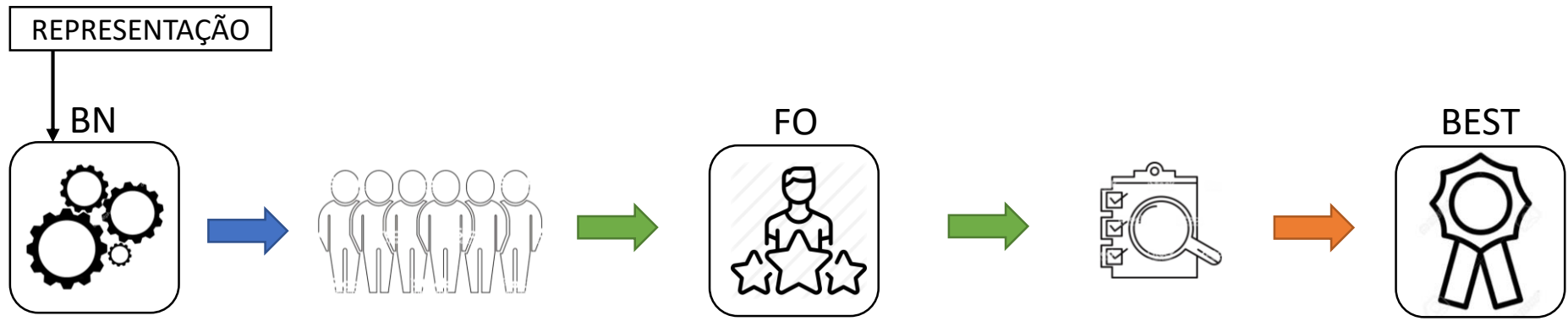
Torneio iterativo



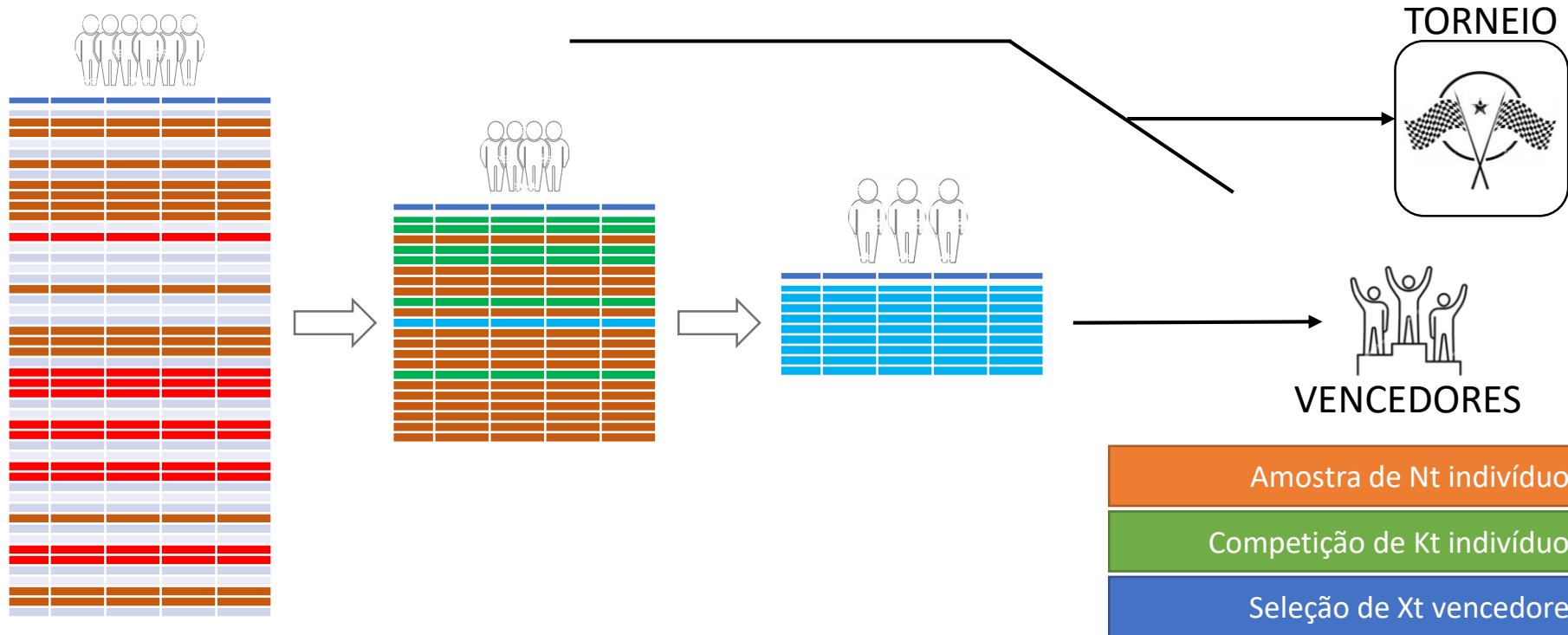
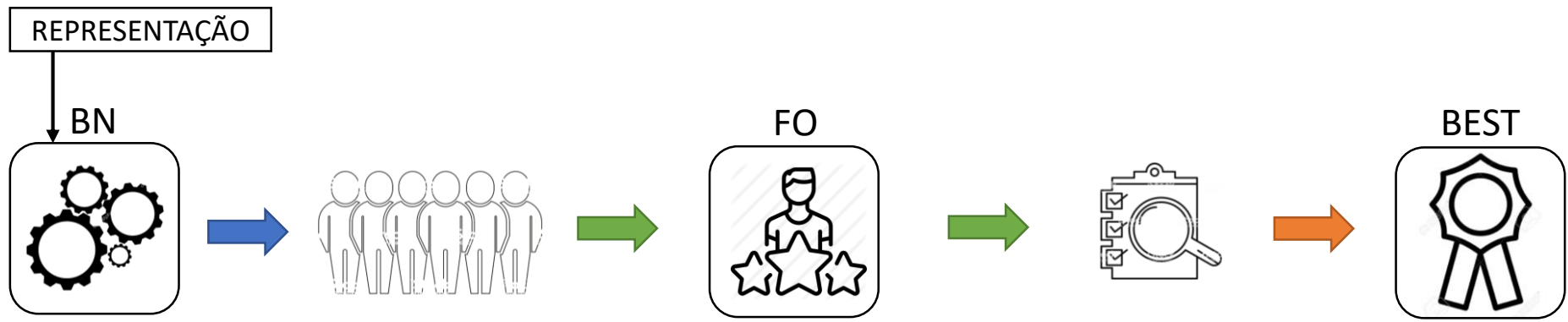
Torneio iterativo



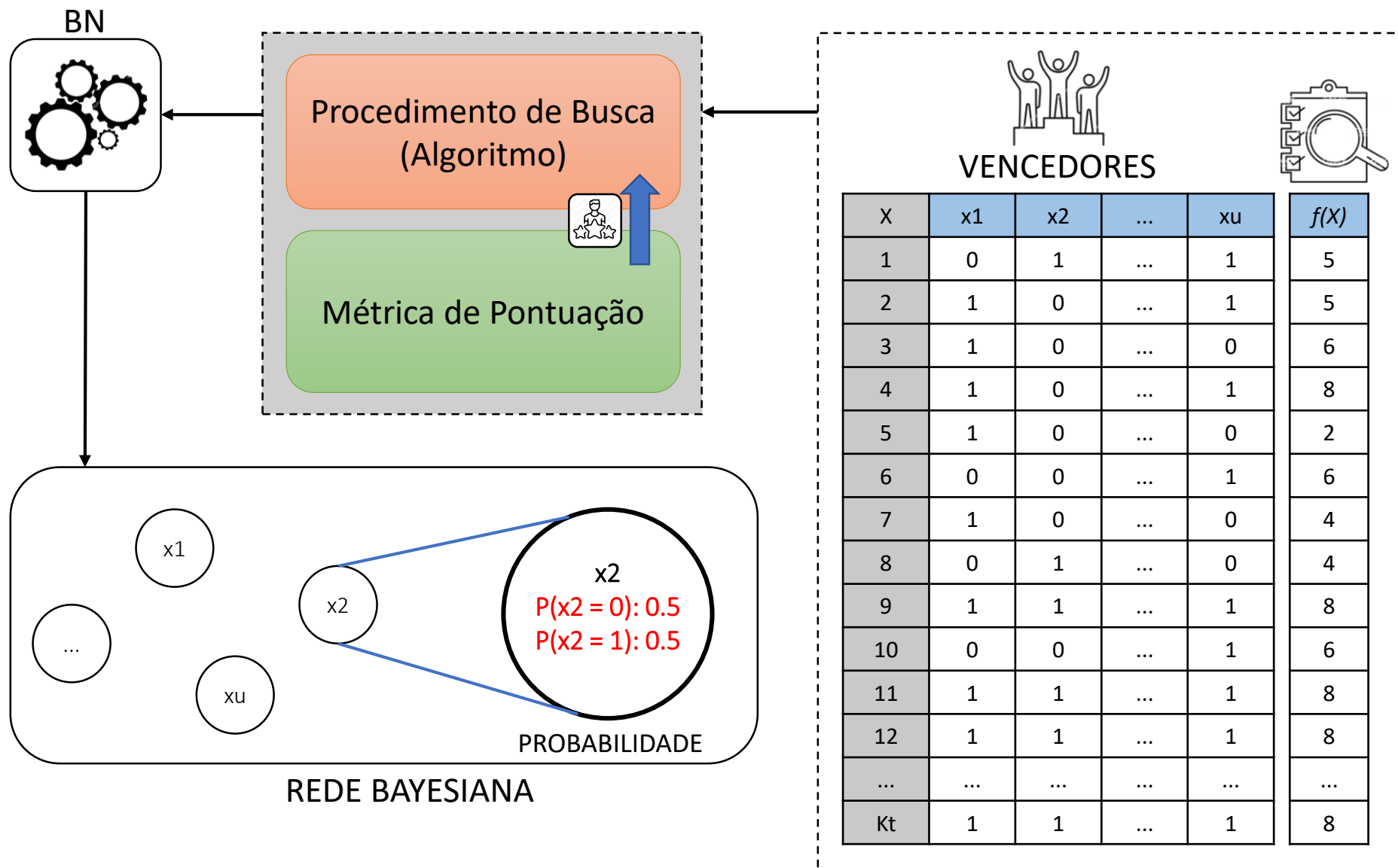
Torneio iterativo



Torneio iterativo



Atualização da Rede Bayesiana



Procedimentos de Busca e Métricas de Pontuação

- Procedimentos de Busca

- *Hill Climbing*
- Busca Tabu

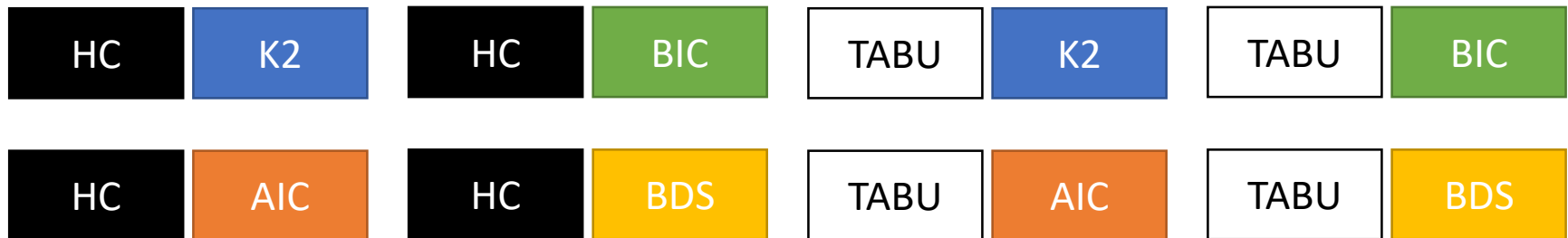
Implementado
em R (3.53)

- Métrica de Pontuação

- K2
- *Akaike Information Criteria* (AIC)
- *Bayesian Information Criteria* (BIC)
- *Bayesian Dirichlet equivalente* (BDS)

Pacote *bnlearn*
para R (3.53)

- Cenários



Hill Climbing

- Move-se para o estado vizinho que proporciona a maior melhoria no valor da função objetivo.
- Apesar de existirem as versões estocásticas, neste trabalho foi utilizada a versão gulosa.

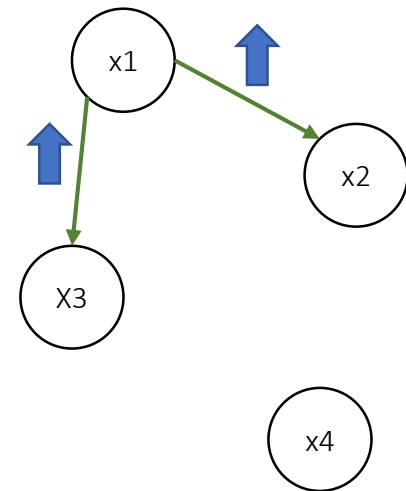
INICIALIZAÇÃO: Faça índice de iterações $t=0$ e escolha estado inicial $x(t) = x(0)$.

ÓTIMO LOCAL: Pare se nenhuma ação (introdução de novo arco) do conjunto G de possíveis ações a partir do estado corrente leva a um estado melhor.

MOVIMENTO: Escolha a ação $g \in G$ que leve ao estado com melhor valor de função objetivo (melhor do que o estado corrente).

ATUALIZAÇÃO: Aplique g sobre o estado corrente $x(t)$ obtendo assim $x(t+1)$.

INCREMENTO: Faça $t = t+1$ e retorne ao passo “ÓTIMO LOCAL”.



Hill Climbing

- Move-se para o estado vizinho que proporciona a maior melhoria no valor da função objetivo.
- Apesar de existirem as versões estocásticas, neste trabalho foi utilizada a versão gulosa.

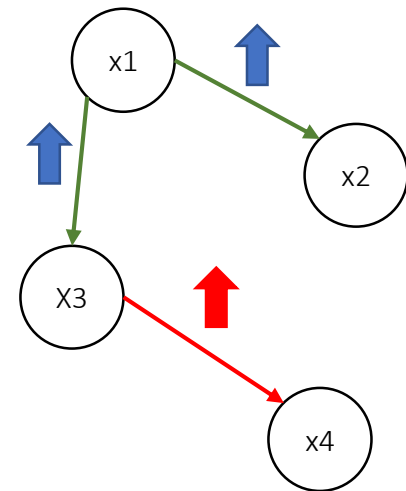
INICIALIZAÇÃO: Faça índice de iterações $t=0$ e escolha estado inicial $x(t) = x(0)$.

ÓTIMO LOCAL: Pare se nenhuma ação (introdução de novo arco) do conjunto G de possíveis ações a partir do estado corrente leva a um estado melhor.

MOVIMENTO: Escolha a ação $g \in G$ que leve ao estado com melhor valor de função objetivo (melhor do que o estado corrente).

ATUALIZAÇÃO: Aplique g sobre o estado corrente $x(t)$ obtendo assim $x(t+1)$.

INCREMENTO: Faça $t = t+1$ e retorne ao passo “ÓTIMO LOCAL”.



Hill Climbing

- Move-se para o estado vizinho que proporciona a maior melhoria no valor da função objetivo.
- Apesar de existirem as versões estocásticas, neste trabalho foi utilizada a versão gulosa.

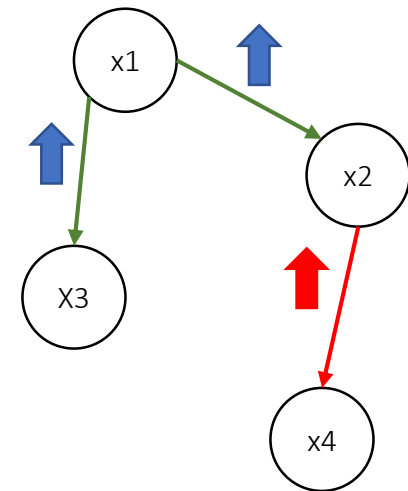
INICIALIZAÇÃO: Faça índice de iterações $t=0$ e escolha estado inicial $x(t) = x(0)$.

ÓTIMO LOCAL: Pare se nenhuma ação (introdução de novo arco) do conjunto G de possíveis ações a partir do estado corrente leva a um estado melhor.

MOVIMENTO: Escolha a ação $g \in G$ que leve ao estado com melhor valor de função objetivo (melhor do que o estado corrente).

ATUALIZAÇÃO: Aplique g sobre o estado corrente $x(t)$ obtendo assim $x(t+1)$.

INCREMENTO: Faça $t = t+1$ e retorne ao passo “ÓTIMO LOCAL”.



Hill Climbing

- Move-se para o estado vizinho que proporciona a maior melhoria no valor da função objetivo.
- Apesar de existirem as versões estocásticas, neste trabalho foi utilizada a versão gulosa.

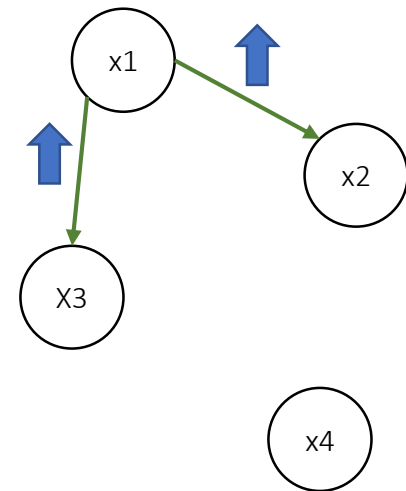
INICIALIZAÇÃO: Faça índice de iterações $t=0$ e escolha estado inicial $x(t) = x(0)$.

ÓTIMO LOCAL: Pare se nenhuma ação (introdução de novo arco) do conjunto G de possíveis ações a partir do estado corrente leva a um estado melhor.

MOVIMENTO: Escolha a ação $g \in G$ que leve ao estado com melhor valor de função objetivo (melhor do que o estado corrente).

ATUALIZAÇÃO: Aplique g sobre o estado corrente $x(t)$ obtendo assim $x(t+1)$.

INCREMENTO: Faça $t = t+1$ e retorne ao passo “ÓTIMO LOCAL”.



Busca Tabu

- Capaz de escapar de ótimos locais permitindo ações que pioram a função objetivo.
- Proibir por um dado “tempo” uma ação já realizada evitando que a solução volte para um estado anterior.

INICIALIZAÇÃO: Faça $t = 0$ e escolha um estado inicial $x(t) = x(0)$ e um limite de iterações $t(\max)$. Crie uma lista (vazia) para ações tabu.

LIMITE DE PARADA: Pare se $t = t(\max)$ e tome como solução o estado com melhor função objetivo dentre os $t(\max)$ estados explorados.

MOVIMENTO: Procure aleatoriamente por uma ação $g \in G$ não tabu que leve a um estado vizinho com função objetivo melhor do que a corrente. Caso não seja encontrada tal ação dentro de um limite de tentativas, selecione a ação tabu menos pior.

ATUALIZAÇÃO: Aplique a ação g sobre o estado corrente $x(t)$ obtendo assim $x(t+1)$.

LISTA TABU: Retire ações que permaneceram na lista por um certo número de iterações e insira ações relacionadas a g .

INCREMENTO: Faça $t = t+1$ e retorne ao passo “Limite de Parada”.

- Cooper, G. F.; Henrskovits, E. A bayesian method for the induction of probabilistic networks from data. Machine Learning, v.9, p. 309-347, 1992.
- Korb K, Nicholson AE (2010). Bayesian Artificial Intelligence. Chapman & Hall/CRC, 2nd edition.

$$P(B_s, D) = P(B_s) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!$$

- Crocomo, Márcio Kassouf. Algoritmo de otimização bayesiano com detecção de comunidades. Tese de Doutorado. Universidade de São Paulo, 2012.
- Ruiz, Carolina. Illustration of the K2 algorithm for learning bayes net structures. Class notes. Department of Computer Science, Worcester Polytechnic Institute, 2016. [disponível em http://web.cs.wpi.edu/~cs539/s11/Projects/k2_algorithm.pdf]

AIC, BIC e BDS

- AIC

- Gujarati, D. N.; Porter, D. C. Econometria Básica. Nova York: McGraw Hill & Bookman, 2008.

$$AIC = e^{2k/n} \frac{SQR}{n}$$

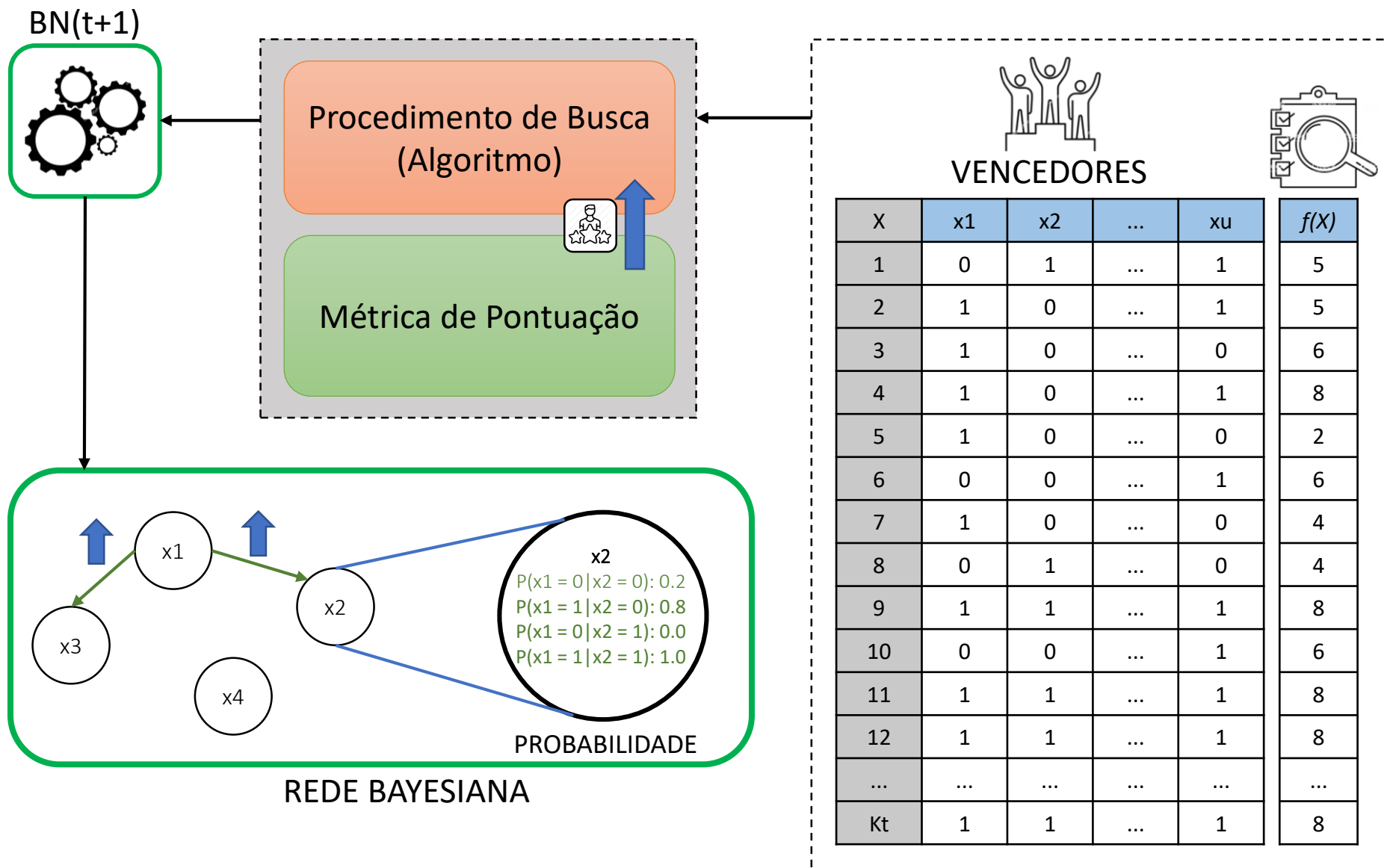
- BIC

- Chickering DM (1995). "A Transformational Characterization of Equivalent Bayesian Network Structures". Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence, 87–98.
- Também chamado de *Schwarz Information Criteria*.
- Similar ao AIC pois também prefere modelos mais simples.

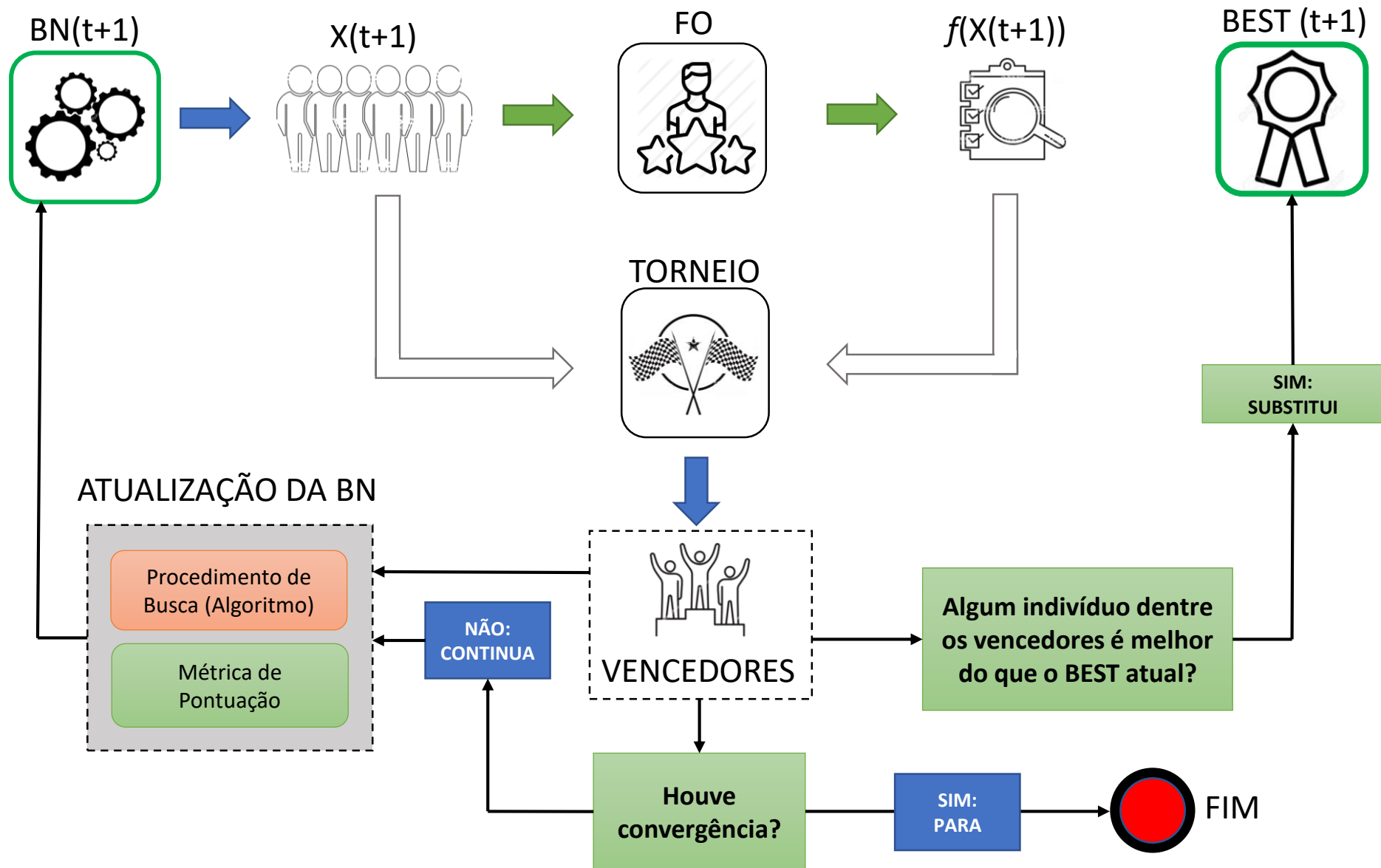
- BDS

- Scutari M (2016). "An Empirical-Bayes Score for Discrete Bayesian Networks". Journal of Machine Learning Research, 52:438–448.
- Métrica que podera a esparsidade da árvore.

Atualização da Rede Bayesiana



Critério de Parada



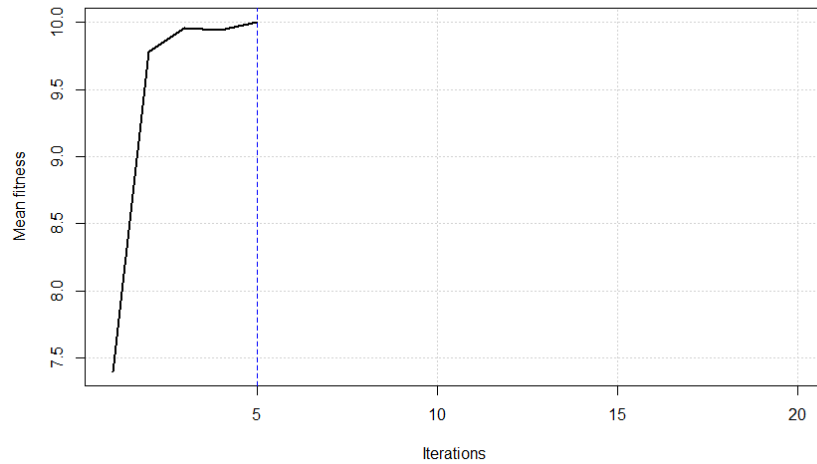
Experimento

- 20 execuções consecutivas, partindo da mesma população inicial para todos os cenários.
- Torneio
 - Amostragem de 50%
 - Torneio com 10% da amostra
- Máximo de 20 gerações
- Tamanho de população igual a 100
- TrapK

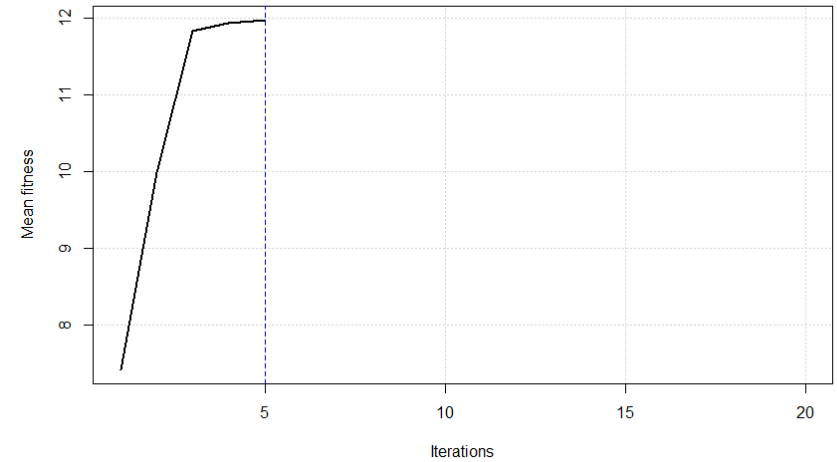
Tamanho do indivíduo	Quantidade de Blocos	Tamanho do bloco
50	5	10
50	10	5
100	10	10
100	25	4
300	30	10

Indivíduo de tamanho 12 e 3 blocos

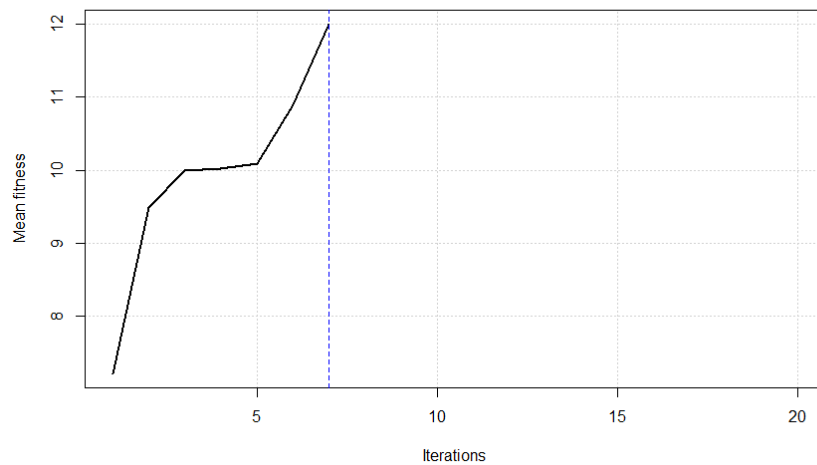
hc-aic: Convergence - Mean fitness



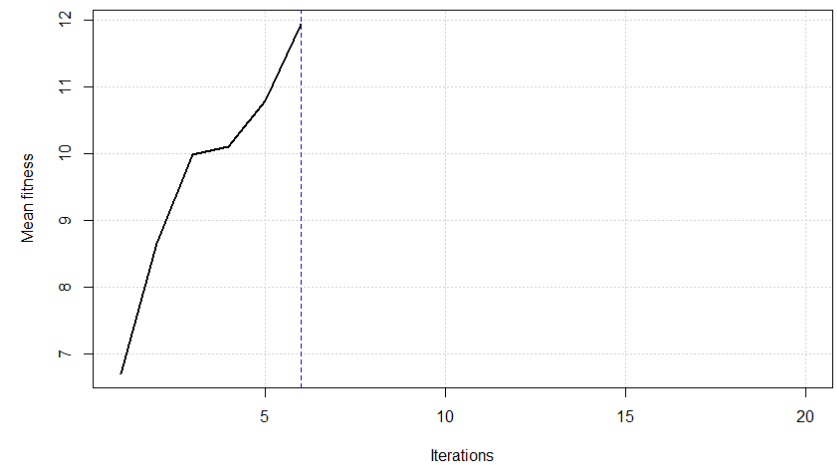
hc-bds: Convergence - Mean fitness



hc-bic: Convergence - Mean fitness

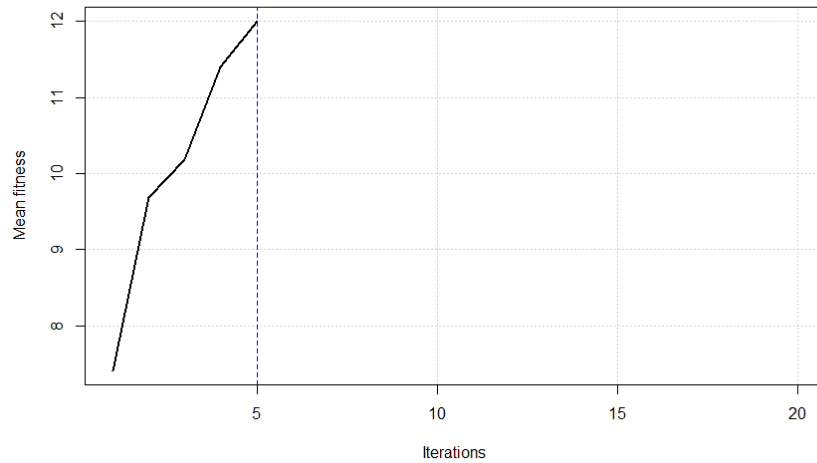


hc-k2: Convergence - Mean fitness

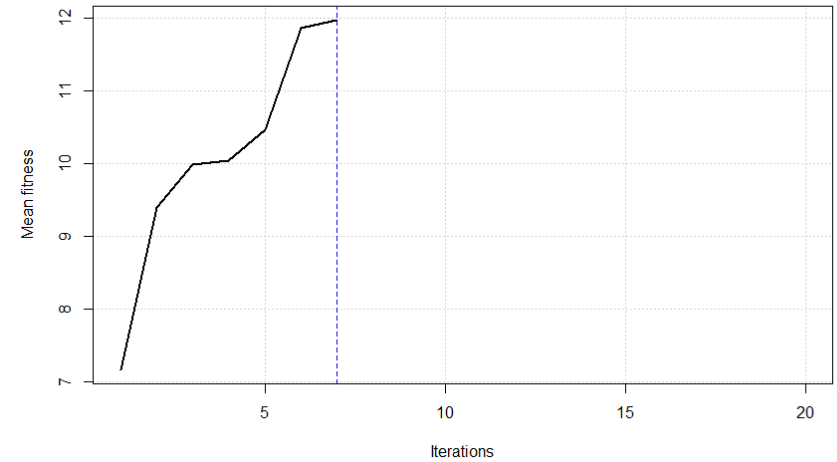


Indivíduo de tamanho 12 e 3 blocos

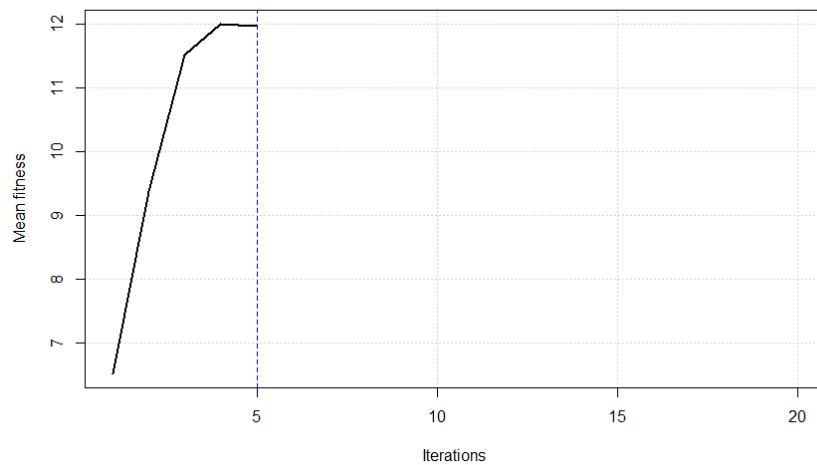
tabu-aic: Convergence - Mean fitness



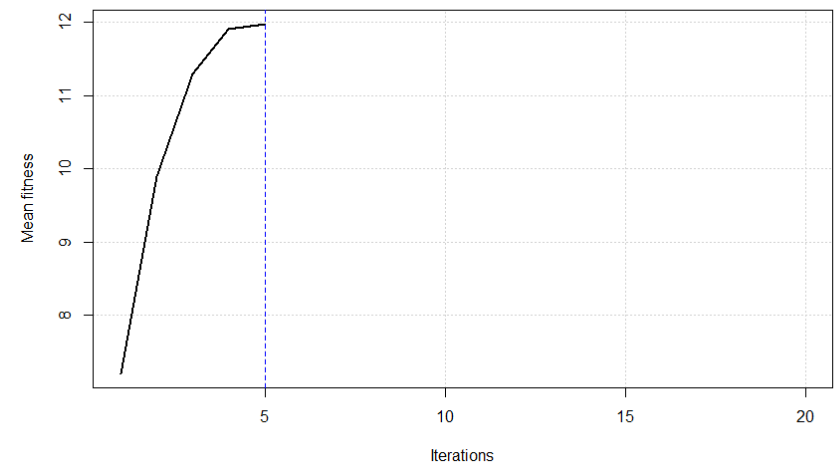
tabu-bds: Convergence - Mean fitness



tabu-bic: Convergence - Mean fitness

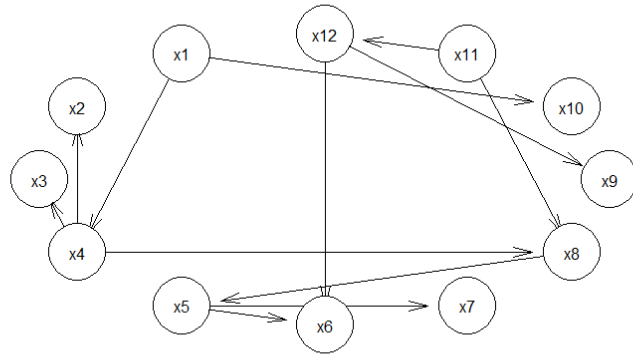


tabu-k2: Convergence - Mean fitness

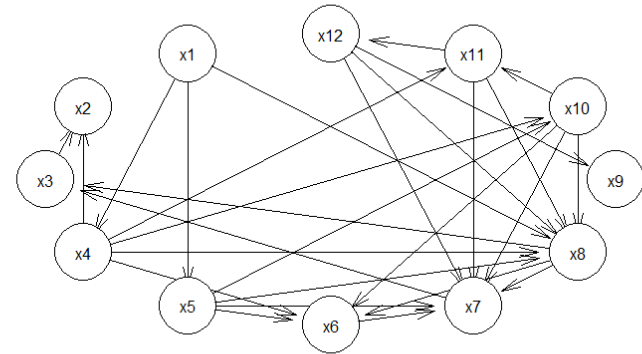


Indivíduo de tamanho 12 e 3 blocos

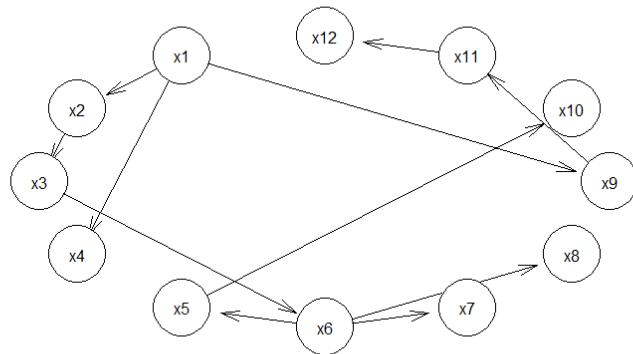
hc-aic



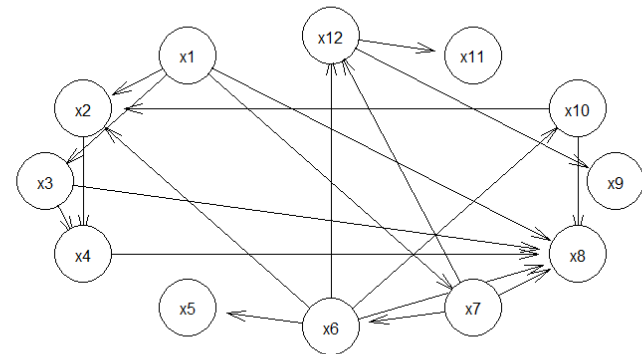
hc-bds



hc-bic

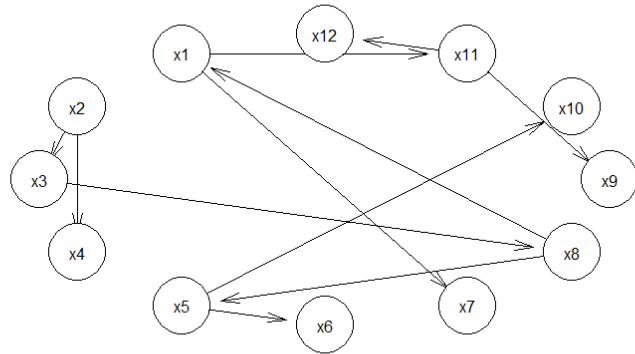


hc-k2

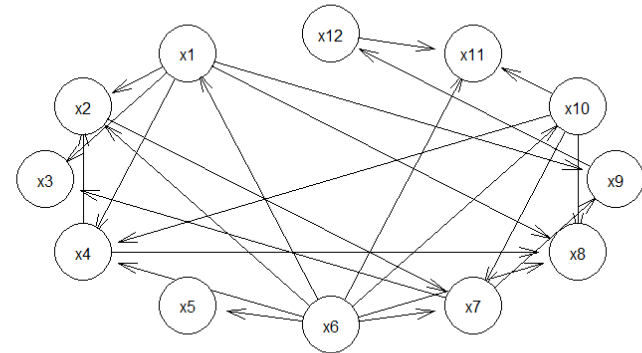


Indivíduo de tamanho 12 e 3 blocos

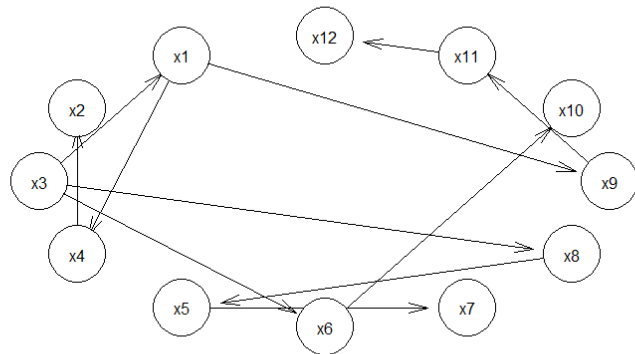
tabu-aic



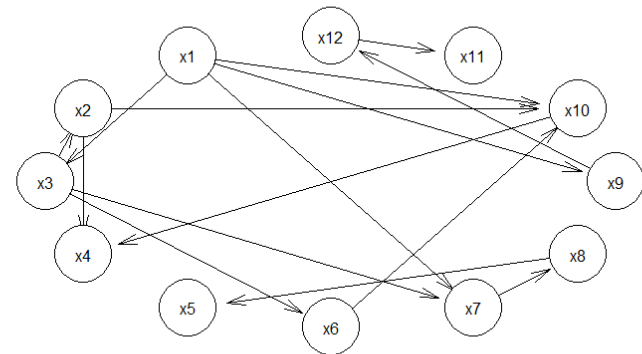
tabu-bds



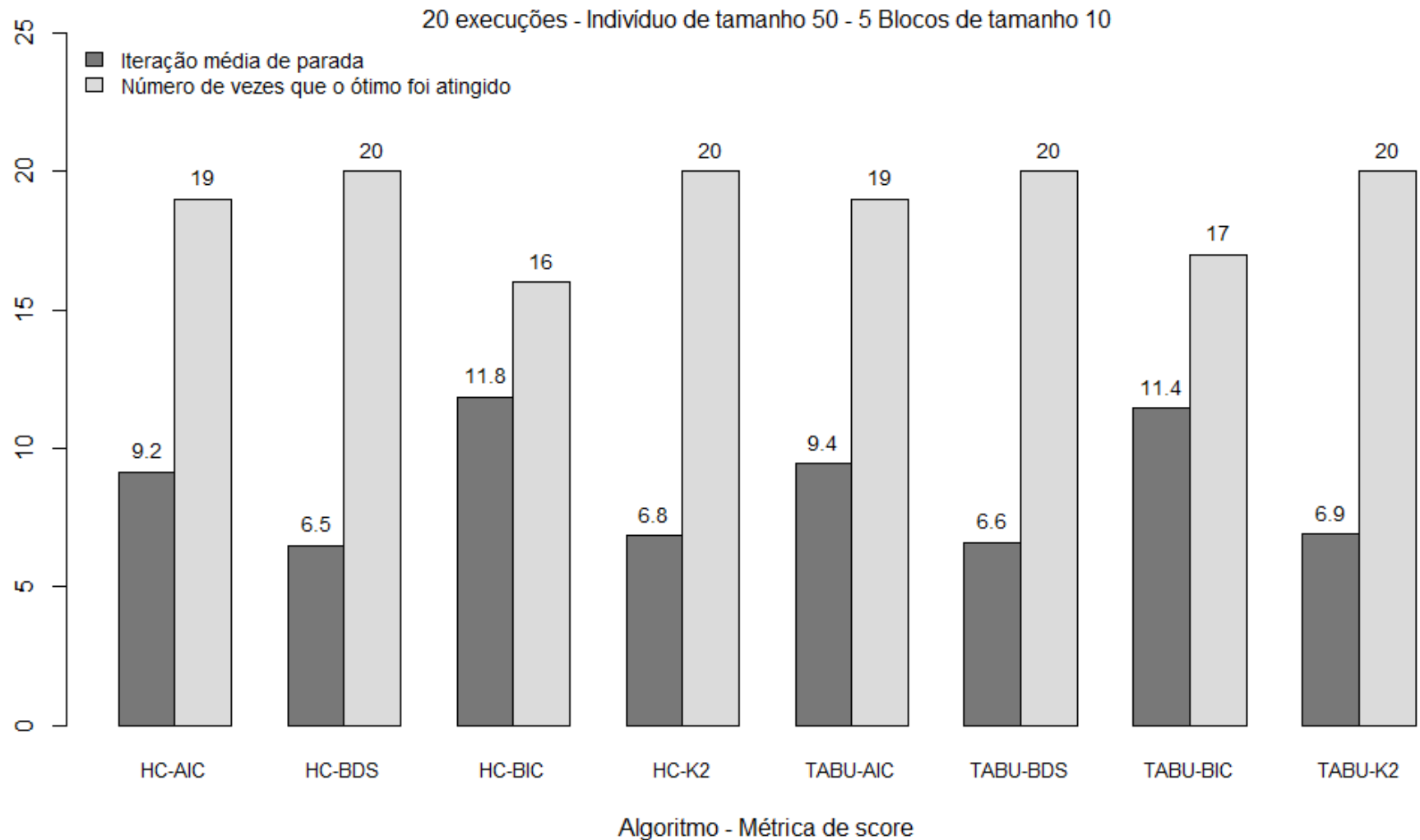
tabu-bic



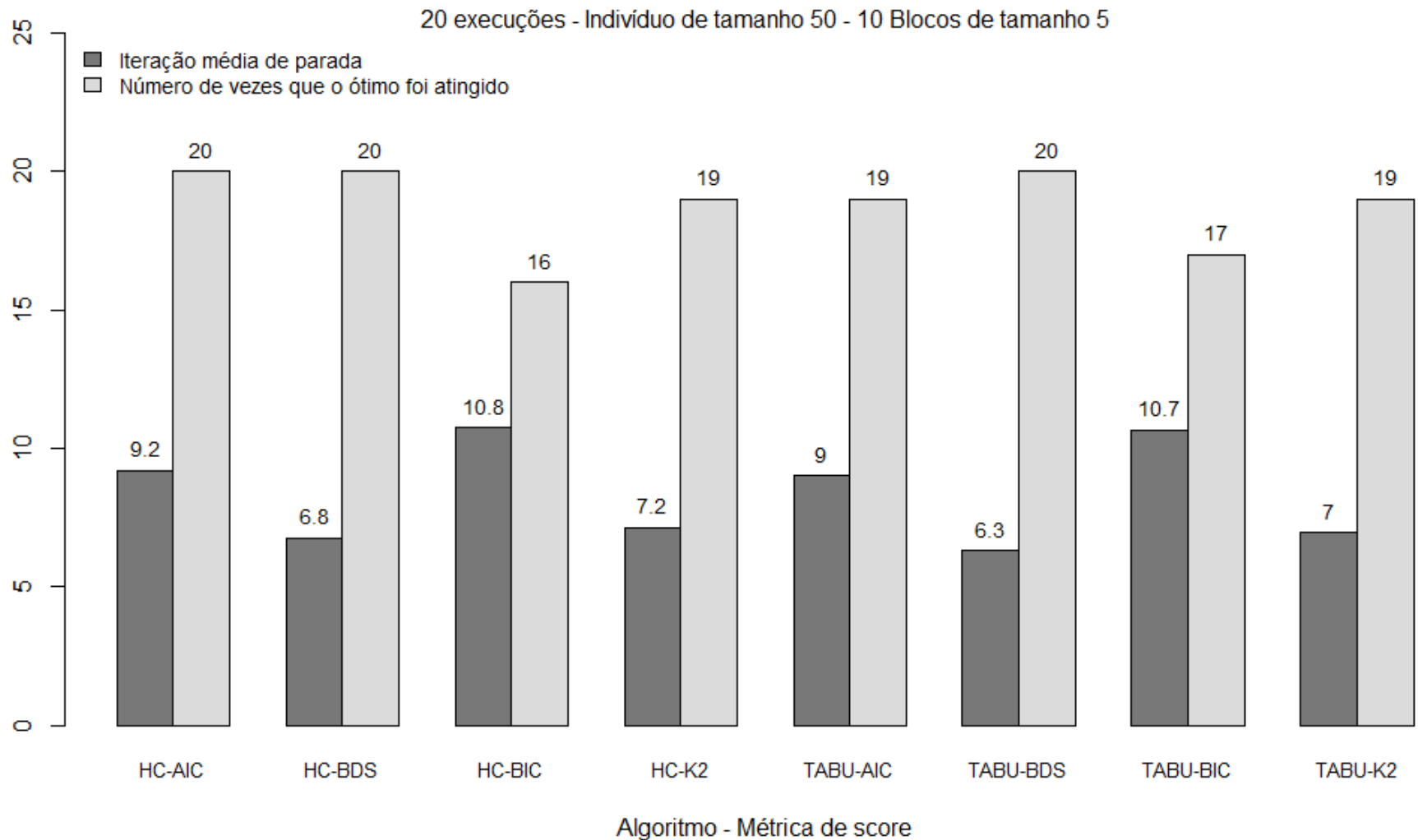
tabu-k2



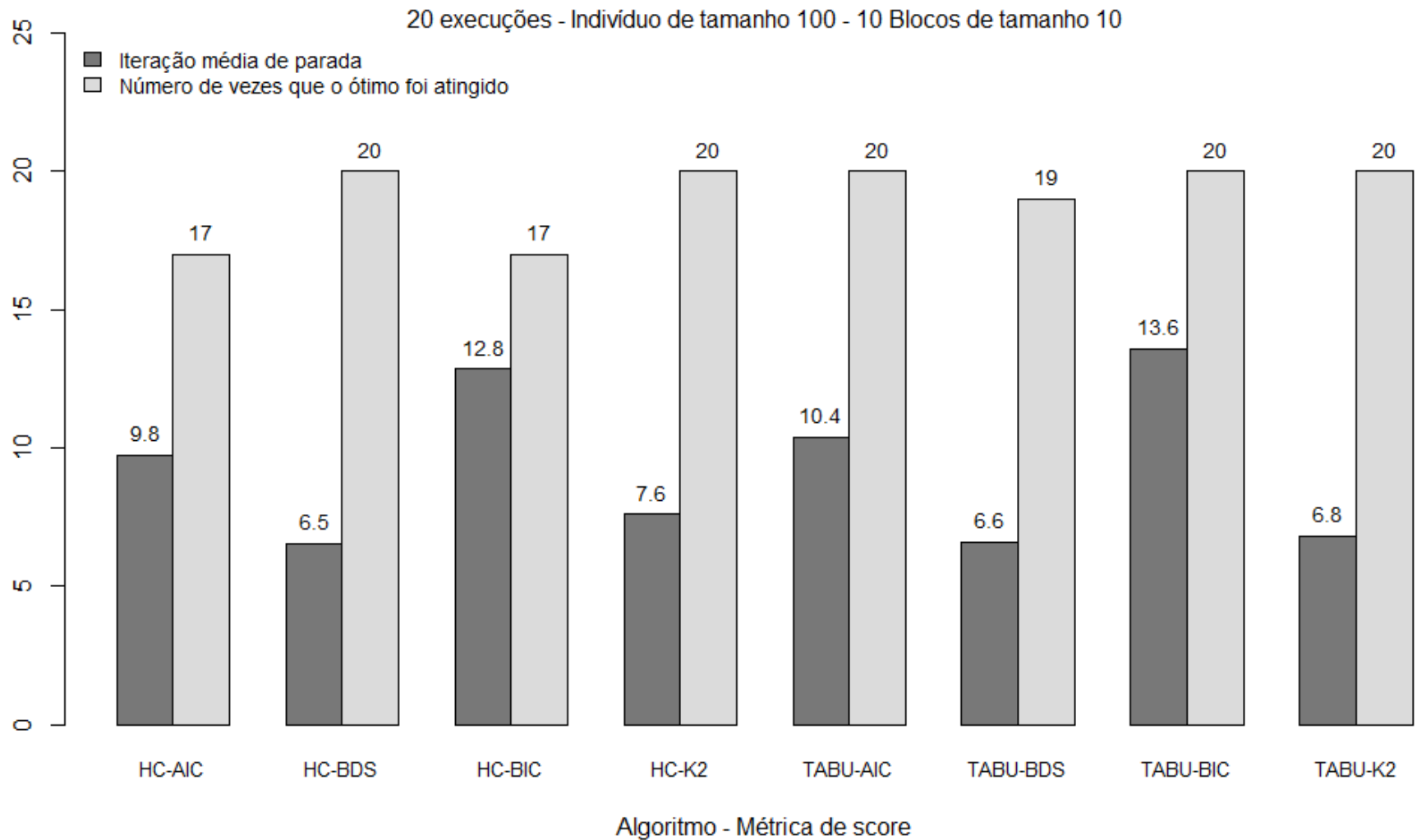
Indivíduo de tamanho 50, 5 blocos de tamanho 10



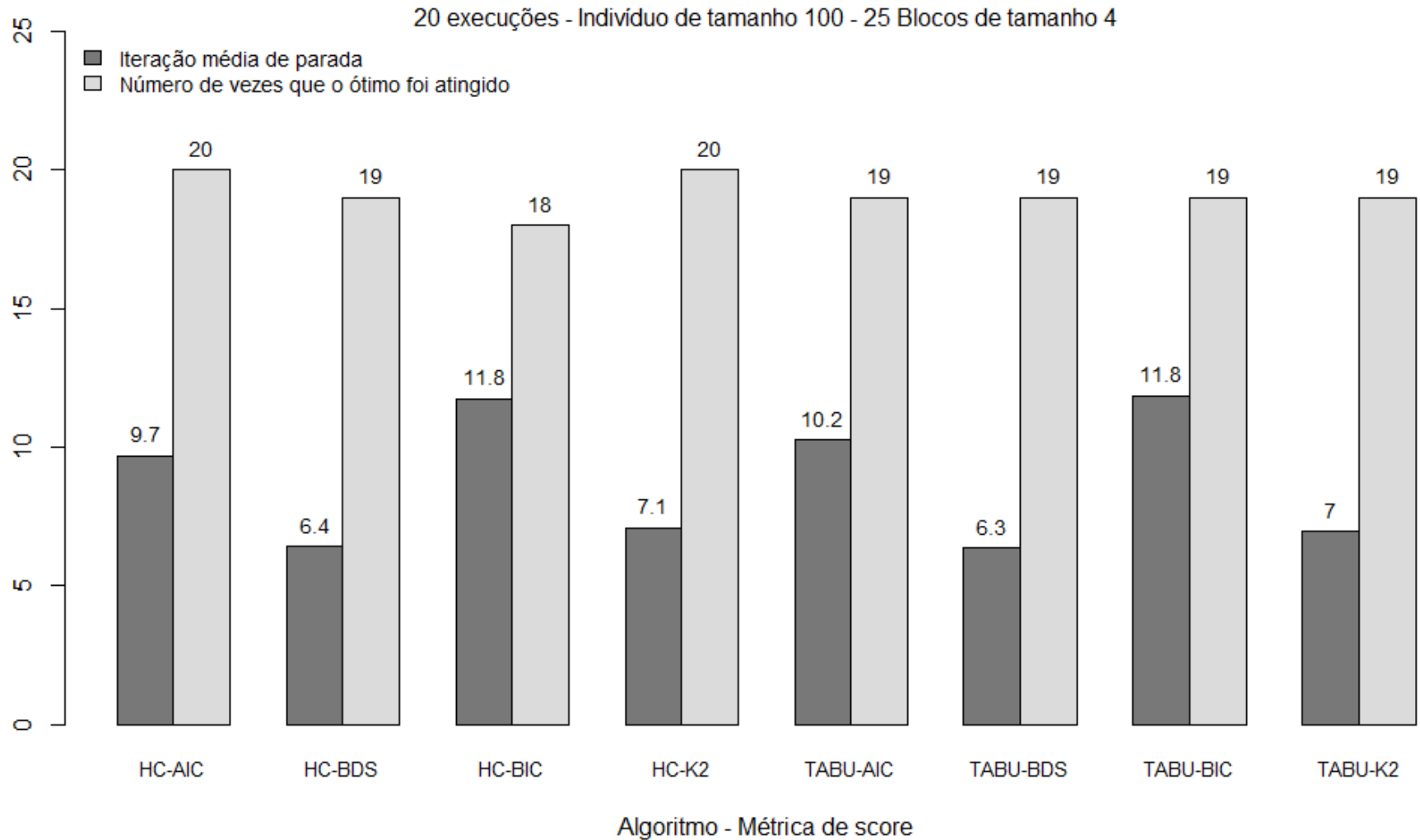
Indivíduo de tamanho 50, 10 blocos de tamanho 5



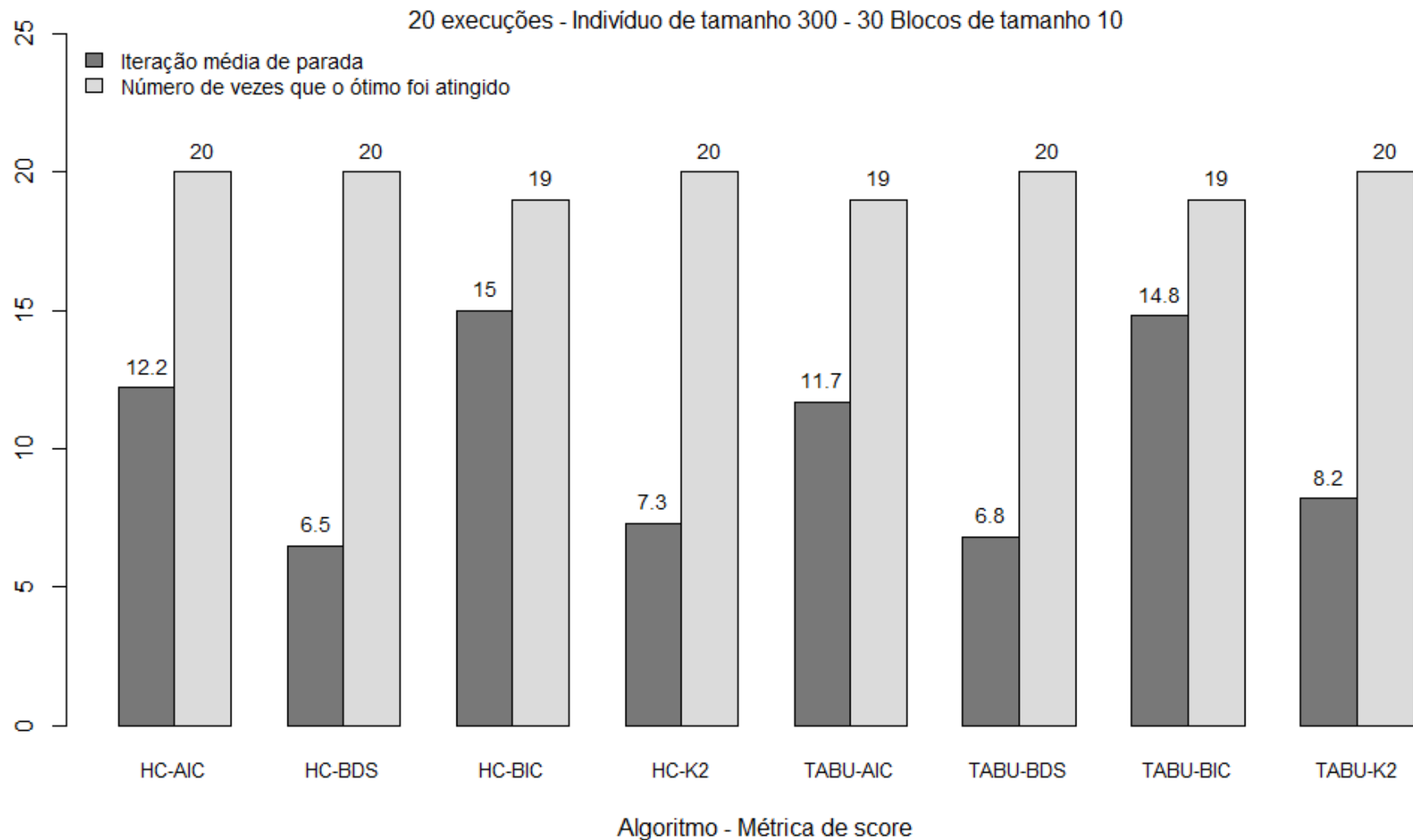
Indivíduo de tamanho 100, 10 blocos de tamanho 10



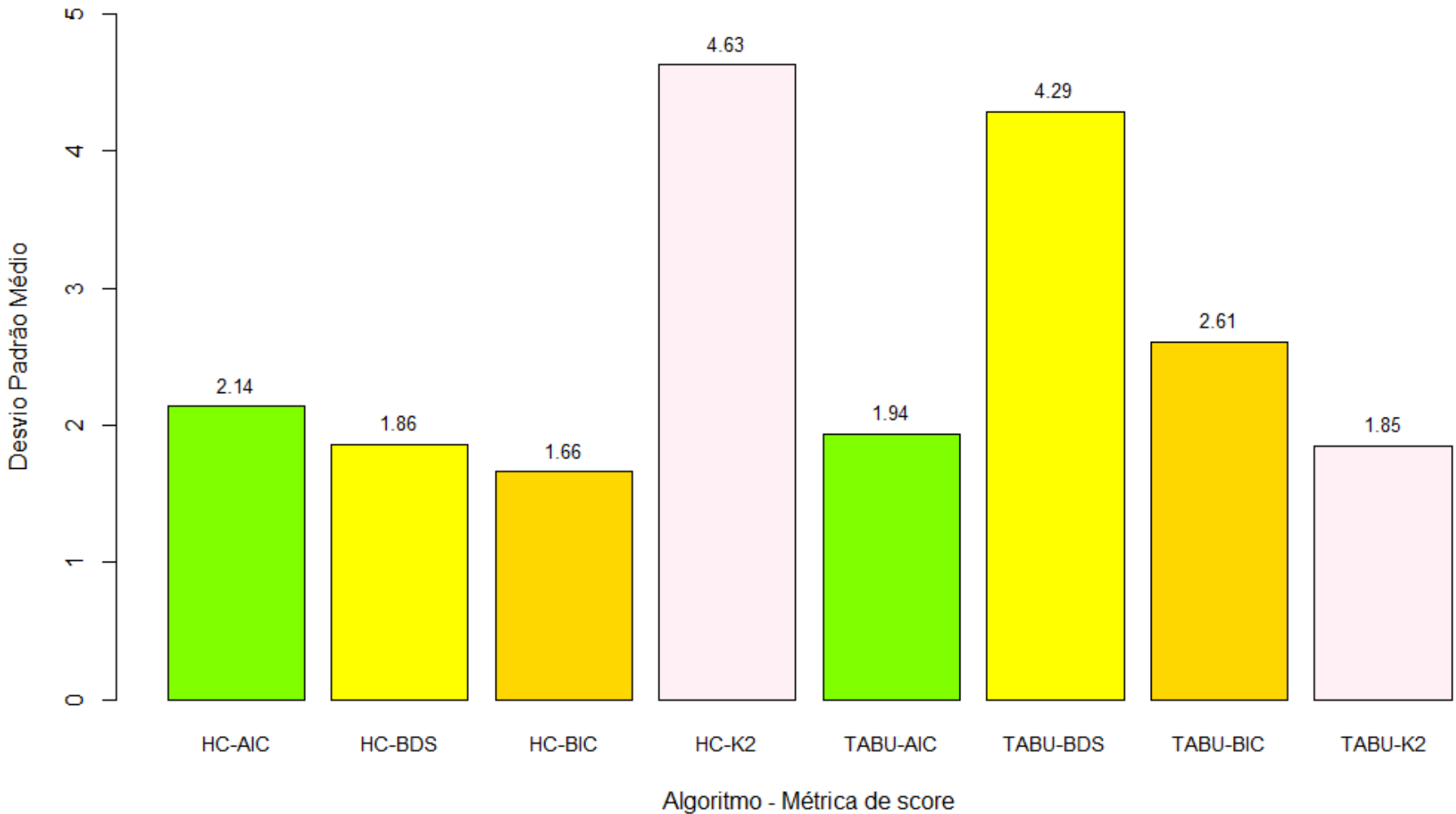
Indivíduo de tamanho 100, 25 blocos de tamanho 4



Indivíduo de tamanho 300, 30 blocos de tamanho 10



Desvio padrão médio (todos os experimentos)



Obrigado!

ciniro@gmail.com
<https://github.com/ciniro>