

ÁRVORE B+

Vanessa Braganholo
Estruturas de Dados e Seus
Algoritmos

ÁRVORES B+

É semelhante à árvore B, exceto por duas características muito importantes:

- Armazena dados somente nas folhas – os nós internos servem apenas de ponteiros
- As folhas são encadeadas

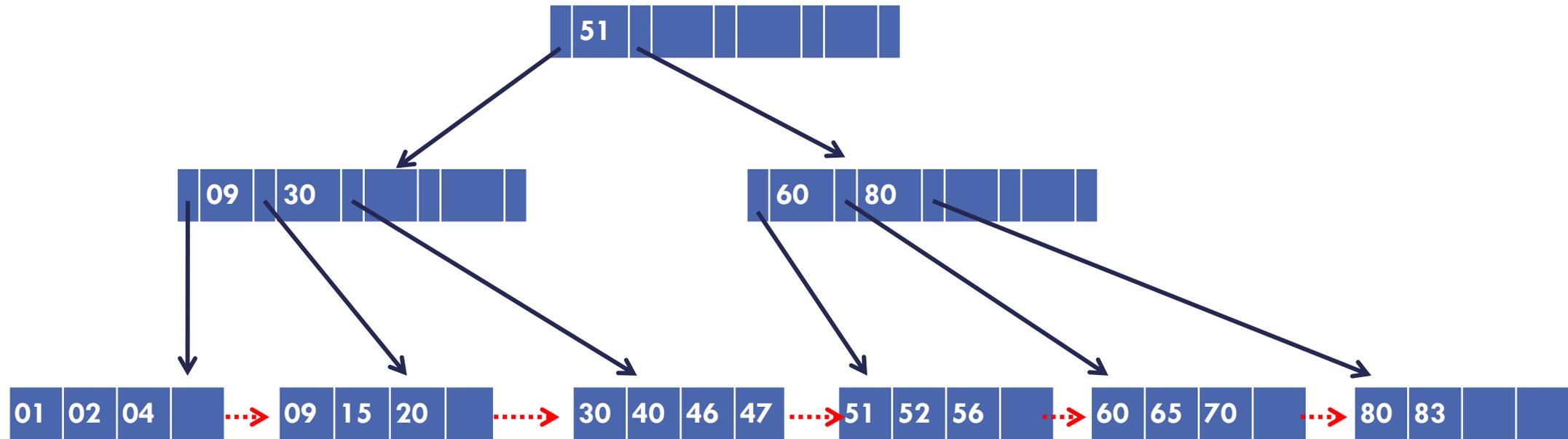
Isso permite o armazenamento dos **dados em um arquivo**, e do **índice em outro arquivo** separado

ÁRVORE B+ NA PRÁTICA

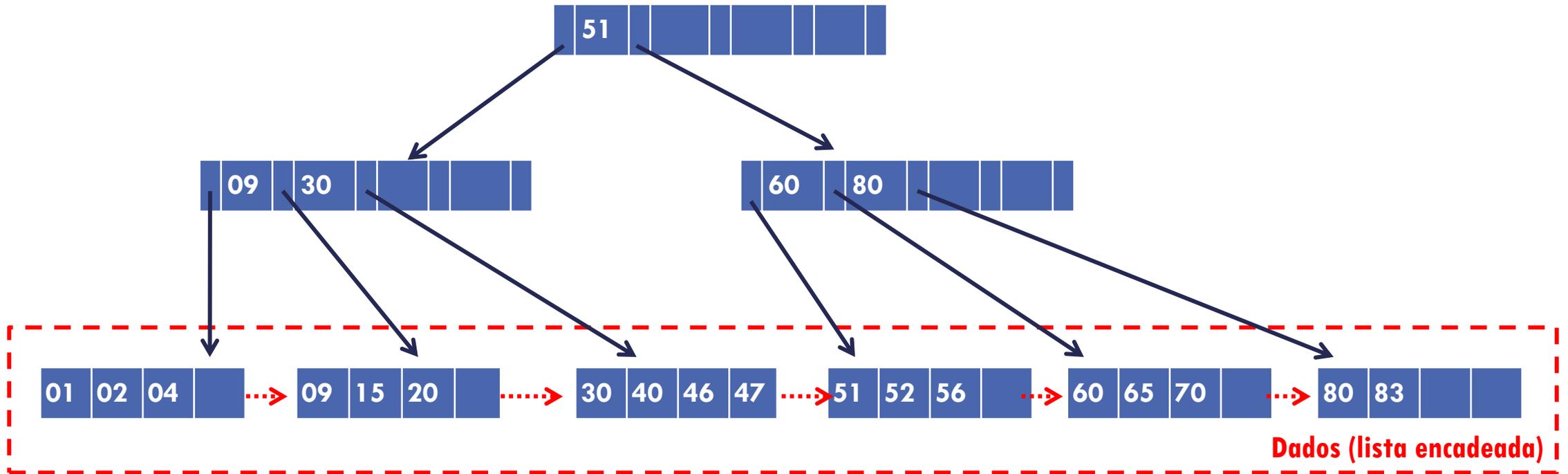
Árvores B+ são muito importantes por sua eficiência, e muito utilizadas na prática:

- Os sistemas de arquivo **NTFS**, **ReiserFS**, **NSS**, **XFS**, e **JFS** utilizam este tipo de árvore para indexação
- Sistemas de Gerência de Banco de Dados como **IBM DB2**, **Informix**, **Microsoft SQL Server**, **Oracle**, **Sybase ASE**, **PostgreSQL**, **Firebird**, **MariaDB** e **SQLite** permitem o uso deste tipo de árvore para indexar tabelas
- Outros sistemas de gerência de dados como o **CouchDB**, **Tokyo Cabinet** e **Tokyo Tyrant** permitem o uso deste tipo de árvore para acesso a dados

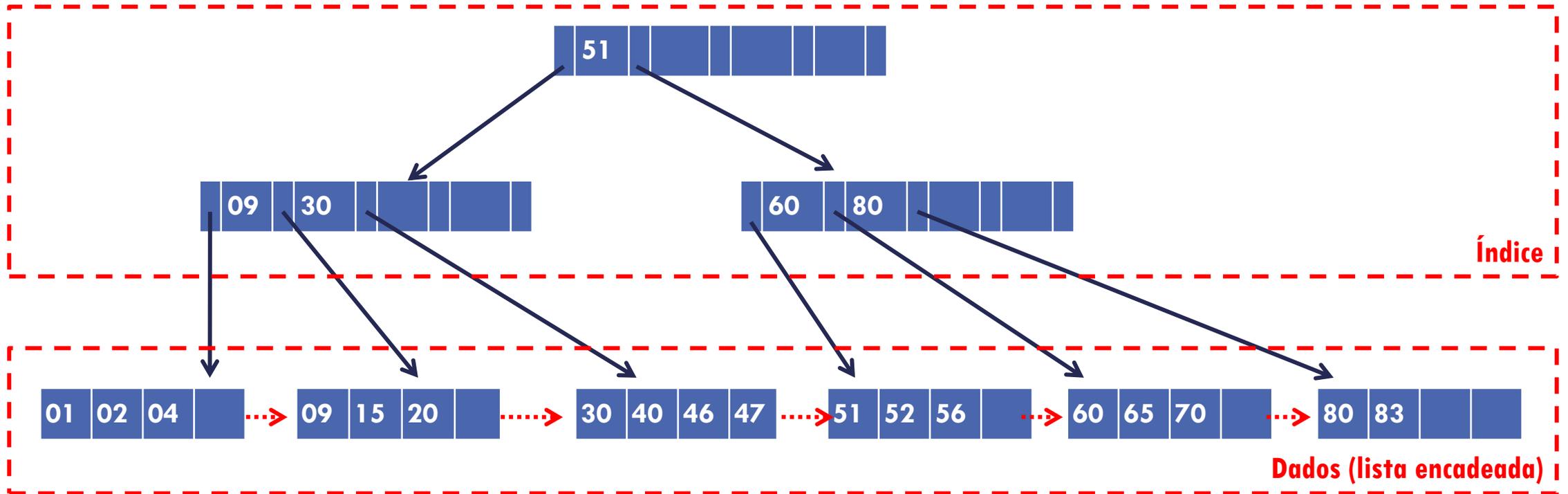
EXEMPLO DE ÁRVORE B+ DE ORDEM D = 2



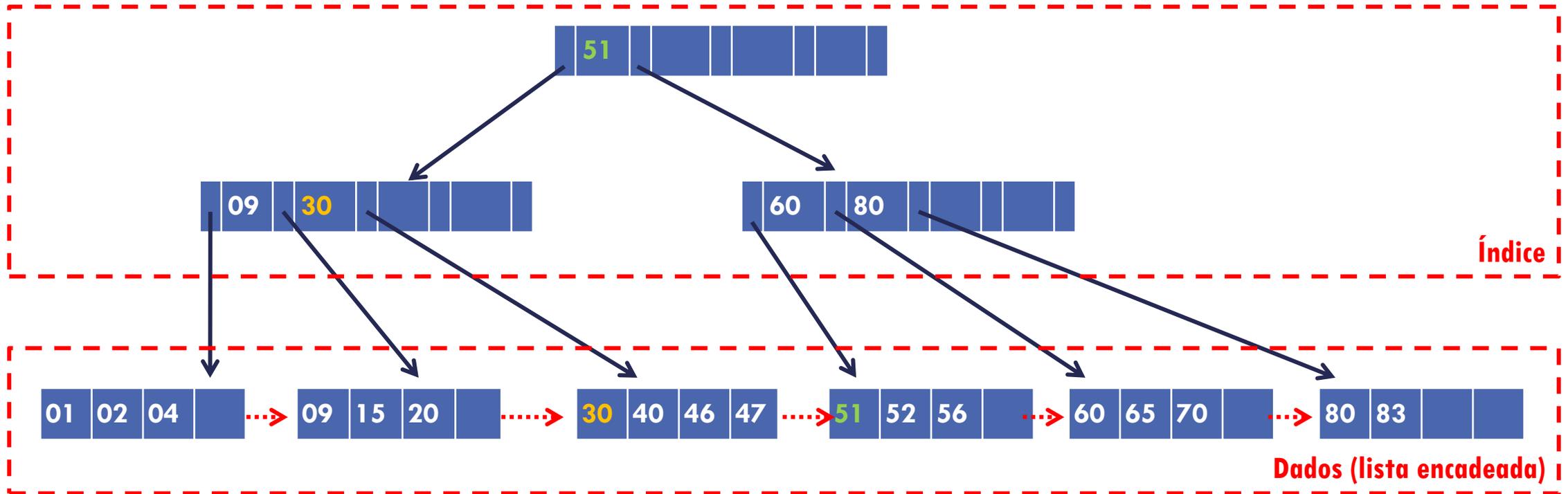
EXEMPLO DE ÁRVORE B+ DE ORDEM D = 2



EXEMPLO DE ÁRVORE B+ DE ORDEM D = 2



EXEMPLO DE ÁRVORE B+ DE ORDEM D = 2



IMPORTANTE:

- Índices **repetem valores** de chave que aparecem nas folhas (diferente do que acontece nas árvores B)

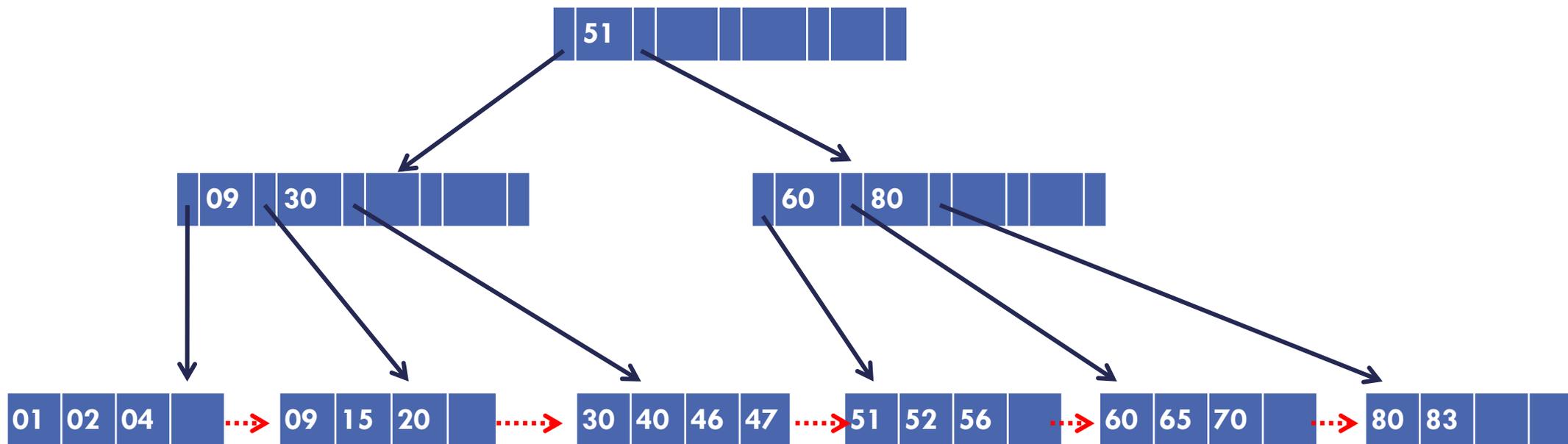
BUSCA

Só se pode ter certeza de que o registro foi encontrado quando se chega em uma folha

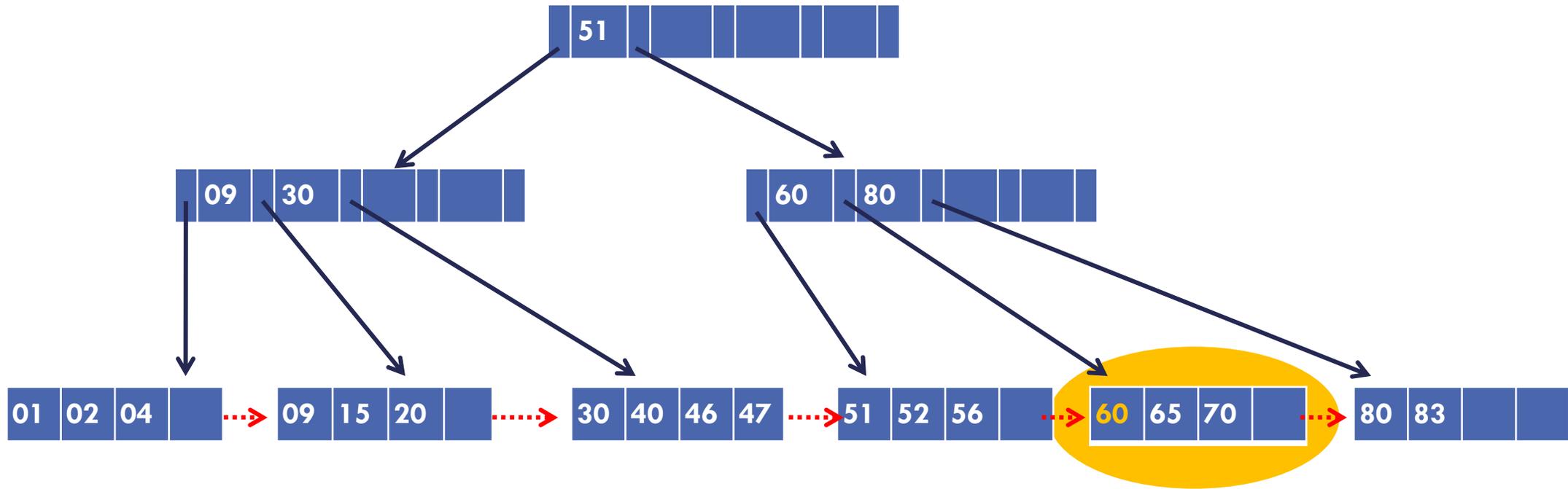
Notar que comparações agora devem considerar a igualdade também

- Achou chave **maior** que a chave buscada, desce pelo ponteiro da esquerda
- Achou chave **igual** à chave buscada ou chegou ao fim da lista de chaves do nó, desce pelo ponteiro da direita

EXEMPLO: BUSCA DE 60



EXEMPLO: BUSCA DE 60



INSERÇÃO

Quando for necessário particionar um nó durante uma inserção, o mesmo raciocínio do particionamento em Árvore B é utilizado

- A diferença é que **para a página pai sobe somente a chave**. O **registro fica na folha**, juntamente com a sua chave

ATENÇÃO:

- Isso vale apenas se o nó que está sendo particionado for uma folha.
- **Se não for folha**, o procedimento é o mesmo utilizado na **árvore B**

EXEMPLO DE INSERÇÃO EM ÁRVORE B+

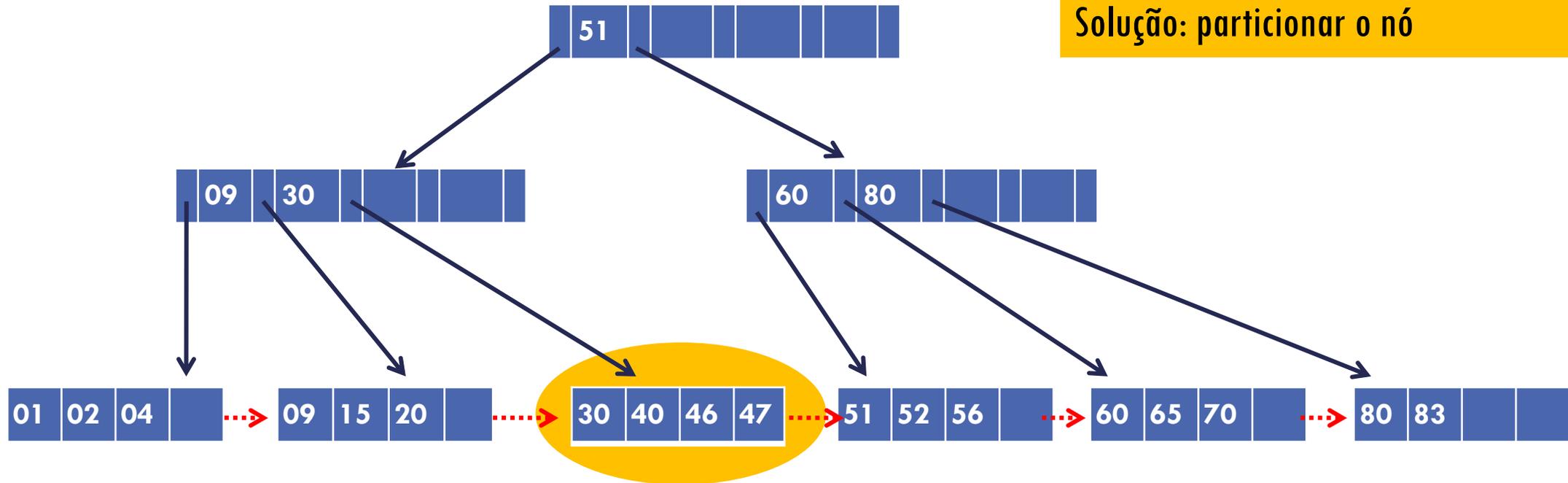
INSERIR CHAVE 32

ordem $d = 2$

Inserir chave 32

Inserção faria página ficar com $2d+1$ chaves

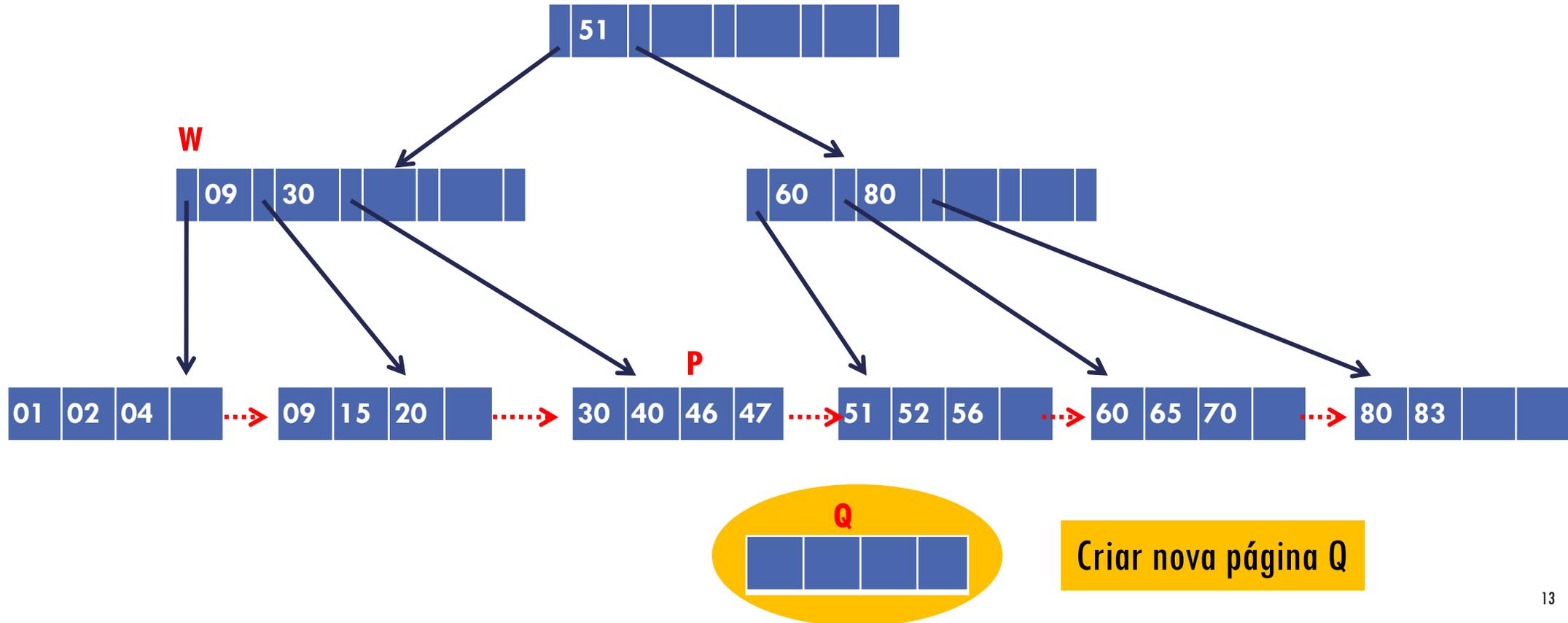
Solução: particionar o nó



EXEMPLO DE INSERÇÃO EM ÁRVORE B+

INSERIR CHAVE 32

ordem $d = 2$



EXEMPLO DE INSERÇÃO EM ÁRVORE B+

INSERIR CHAVE 32

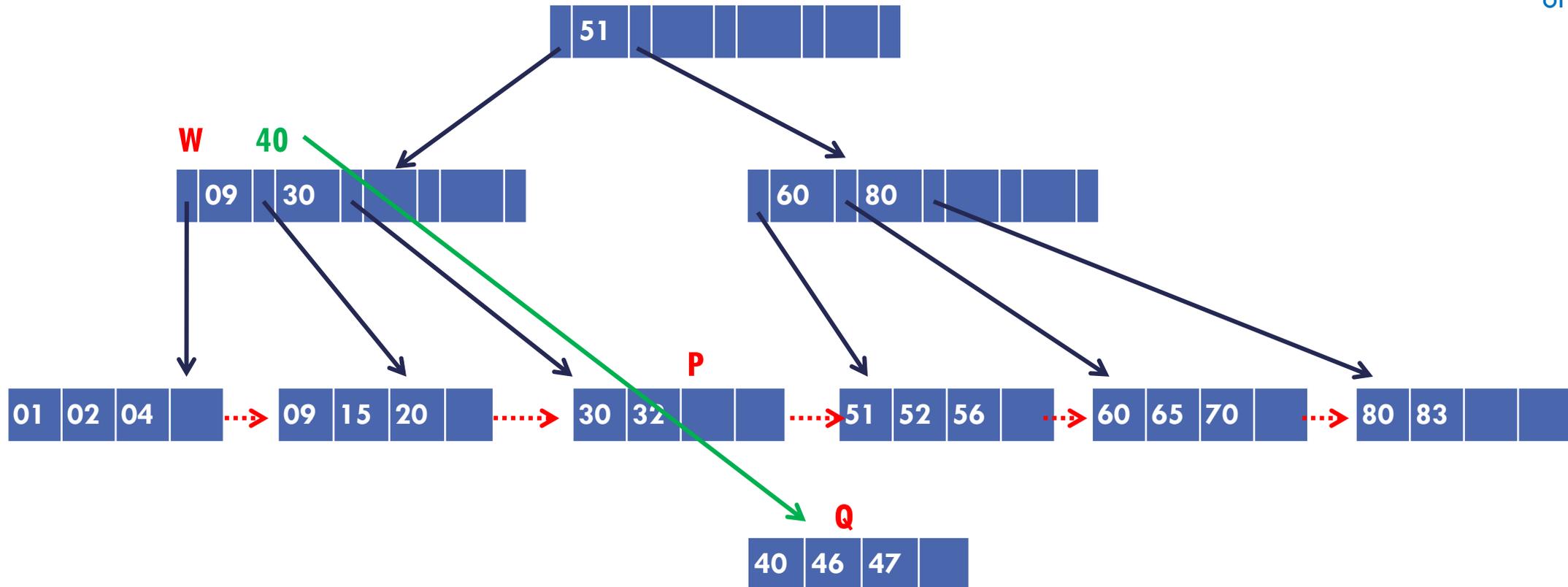
Dividir as chaves entre as duas páginas (30; 32; 40; 46; 47)

d chaves na página original P

chave $d+1$ sobe para nó pai W (mas registro é mantido na nova página)

$d+1$ chaves restantes na nova página Q

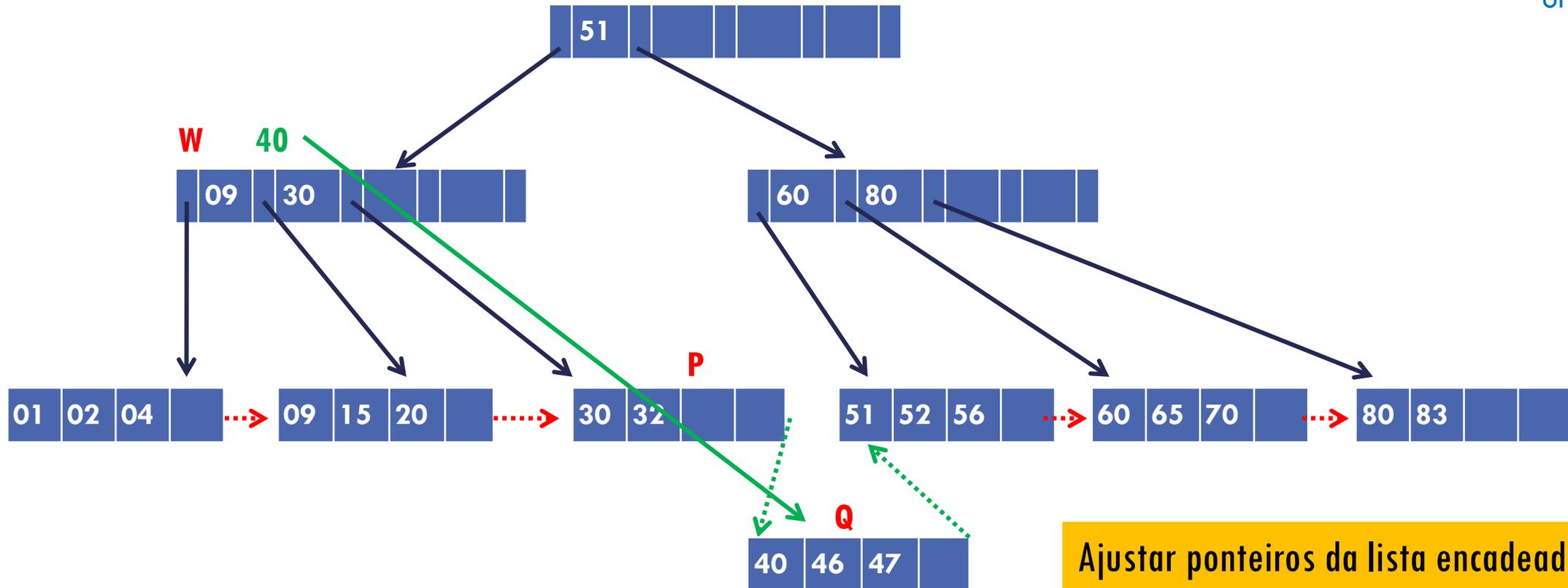
ordem $d = 2$



EXEMPLO DE INSERÇÃO EM ÁRVORE B+

INSERIR CHAVE 32

ordem d = 2

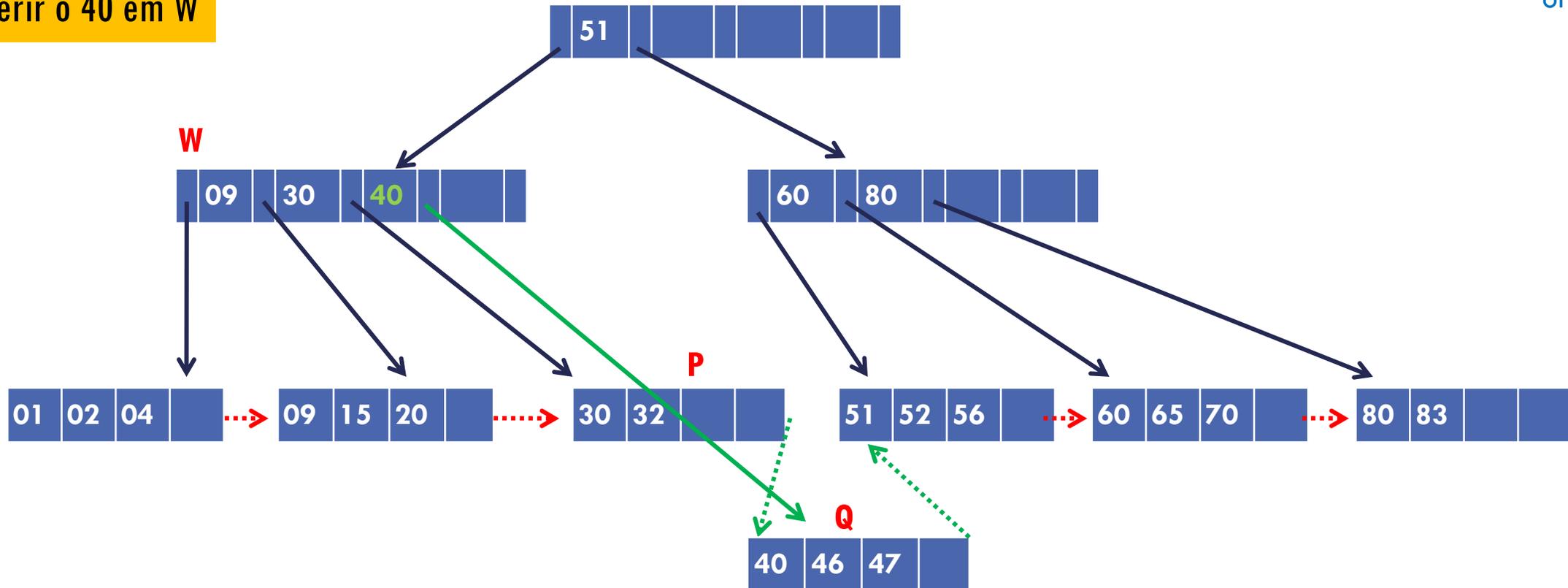


EXEMPLO DE INSERÇÃO EM ÁRVORE B+

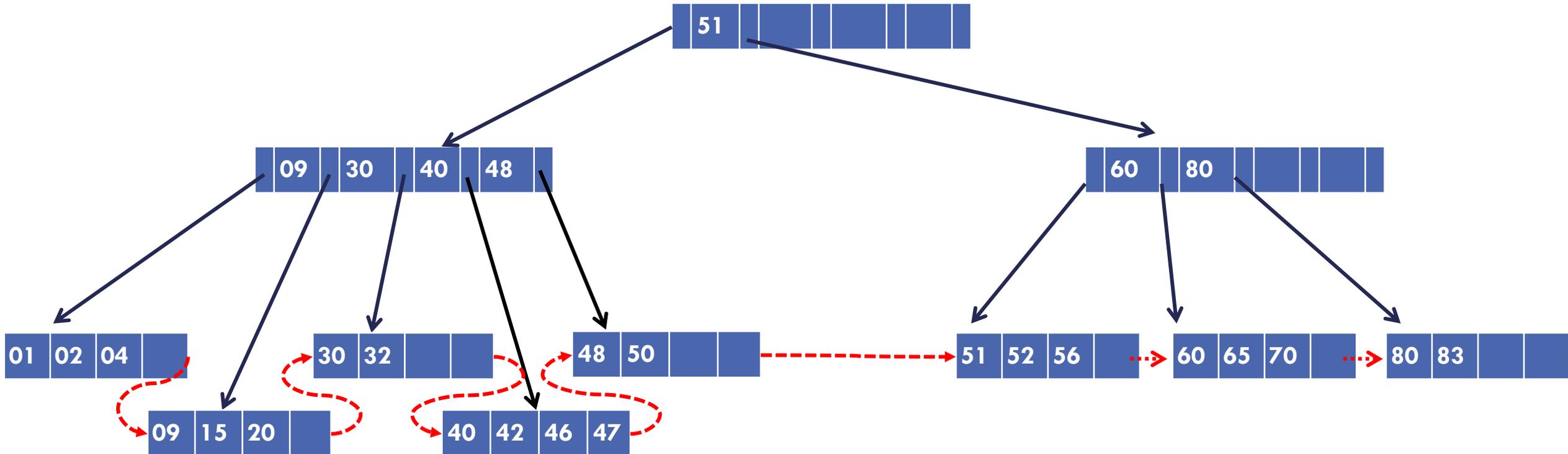
INSERIR CHAVE 32

Inserir o 40 em W

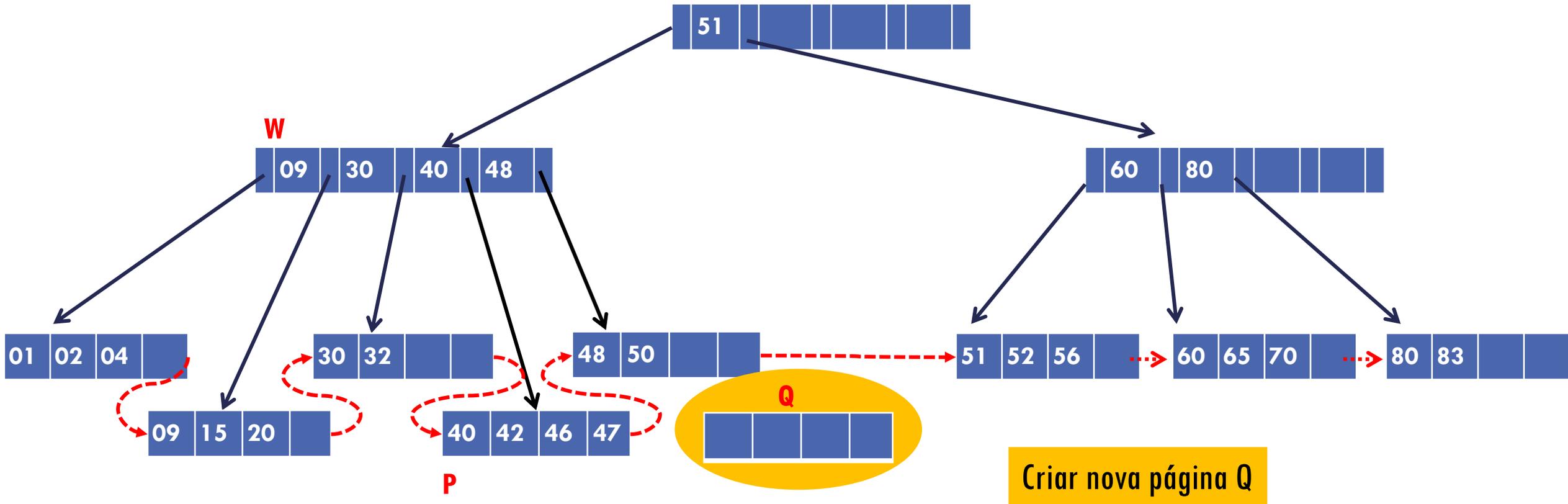
ordem $d = 2$



EXEMPLO DE INSERÇÃO COM PARTICIONAMENTO DE NÓ INTERNO: INSERIR CHAVE 44

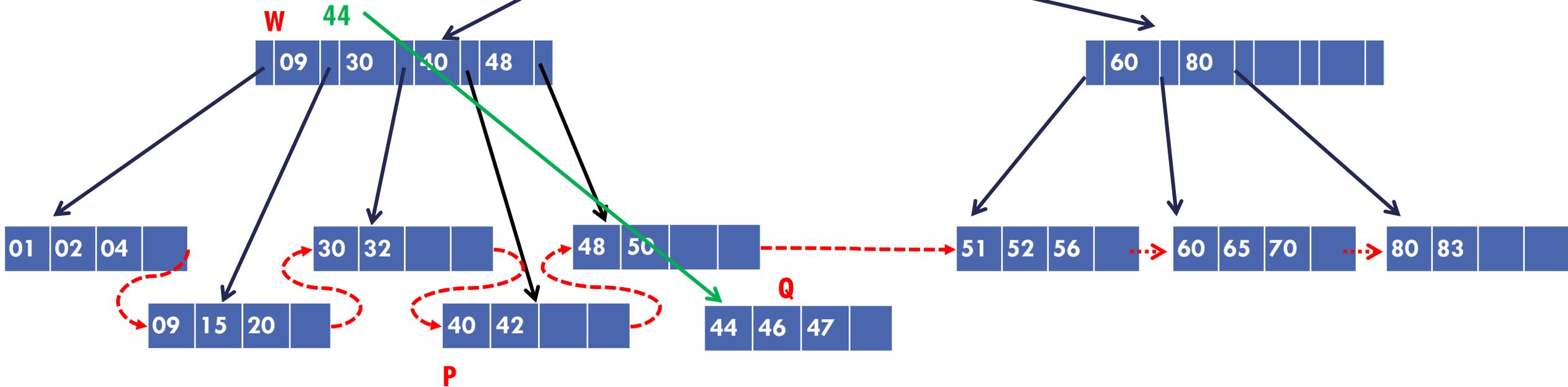


EXEMPLO DE INSERÇÃO COM PARTICIONAMENTO DE NÓ INTERNO: INSERIR CHAVE 44

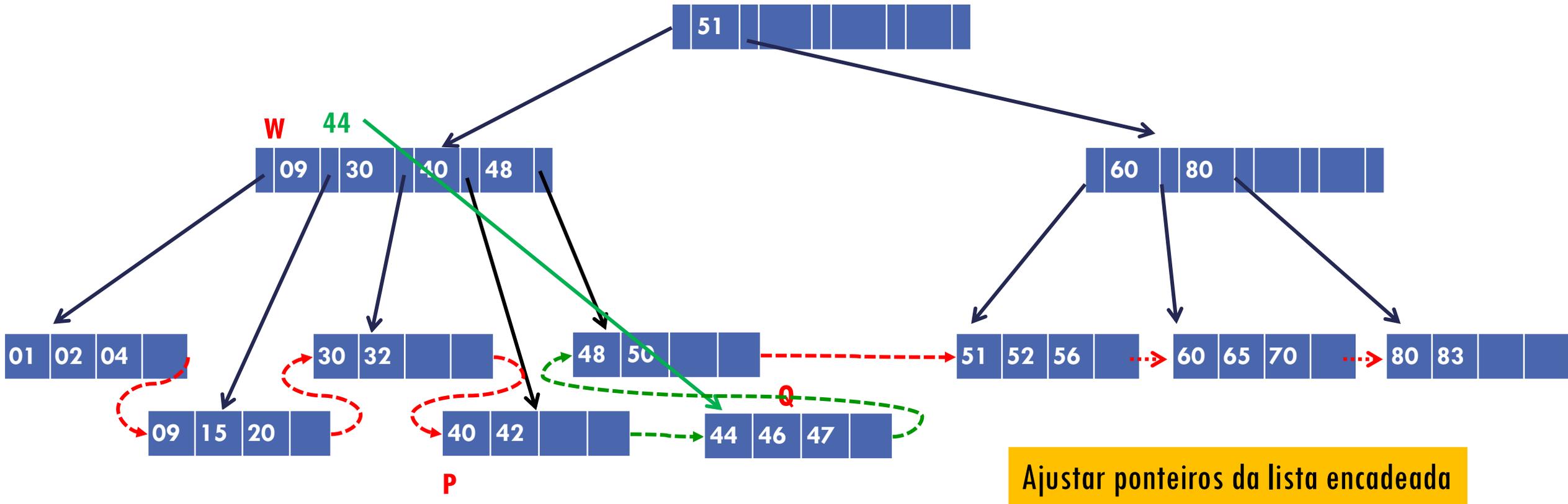


EXEMPLO DE INSERÇÃO COM PARTICIONAMENTO DE NÓ INTERNO: INSERIR CHAVE 44

Dividir as chaves entre as duas páginas (40; 42; 44; 46; 47)
 d chaves na página original P
 chave d+1 sobe para nó pai W (mas registro é mantido na nova página)
 d+1 chaves restantes na nova página Q



EXEMPLO DE INSERÇÃO COM PARTICIONAMENTO DE NÓ INTERNO: INSERIR CHAVE 44

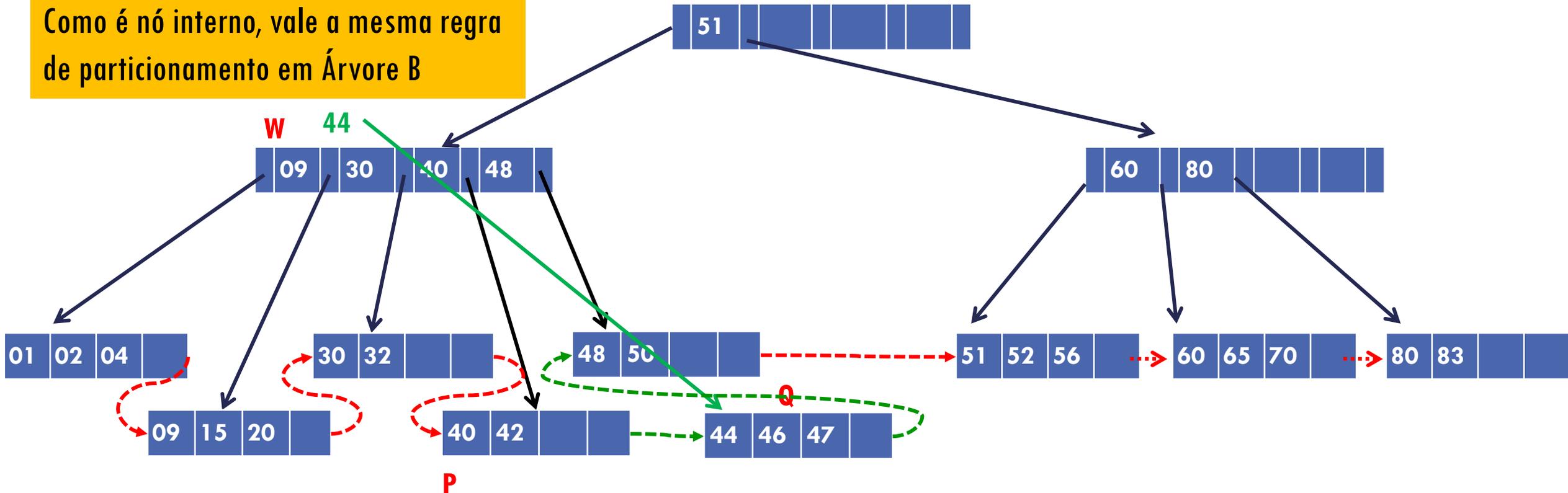


EXEMPLO DE INSERÇÃO COM PARTICIONAMENTO DE NÓ INTERNO: INSERIR CHAVE 44

Inserir 44 em W

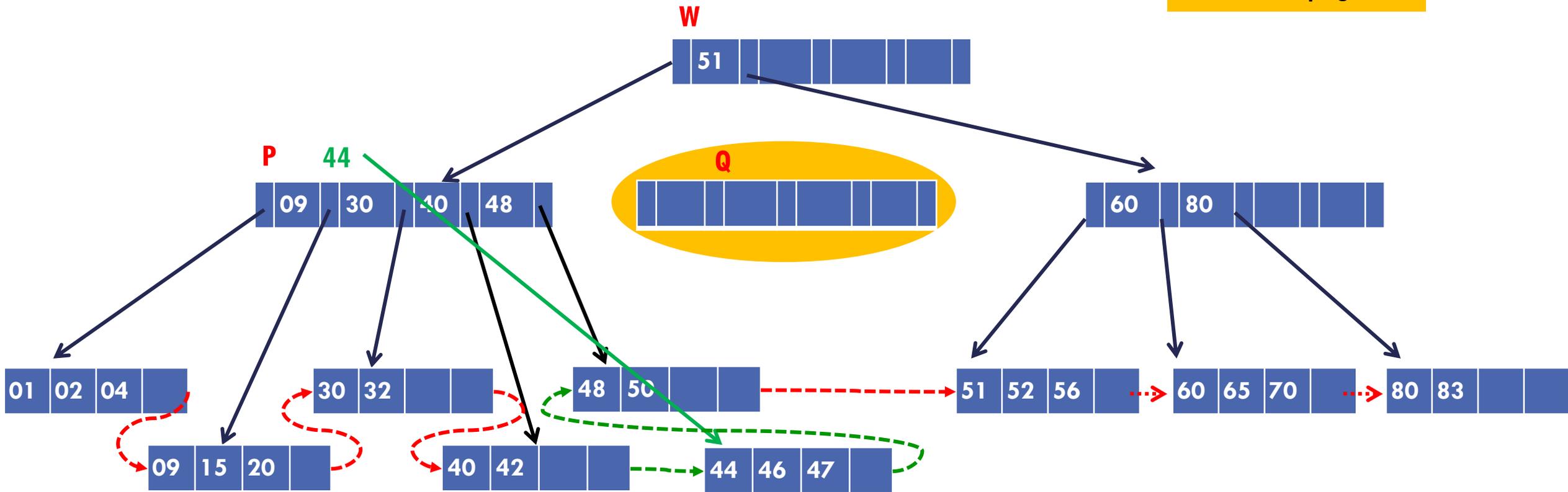
Não há espaço: particionar

Como é nó interno, vale a mesma regra de particionamento em Árvore B



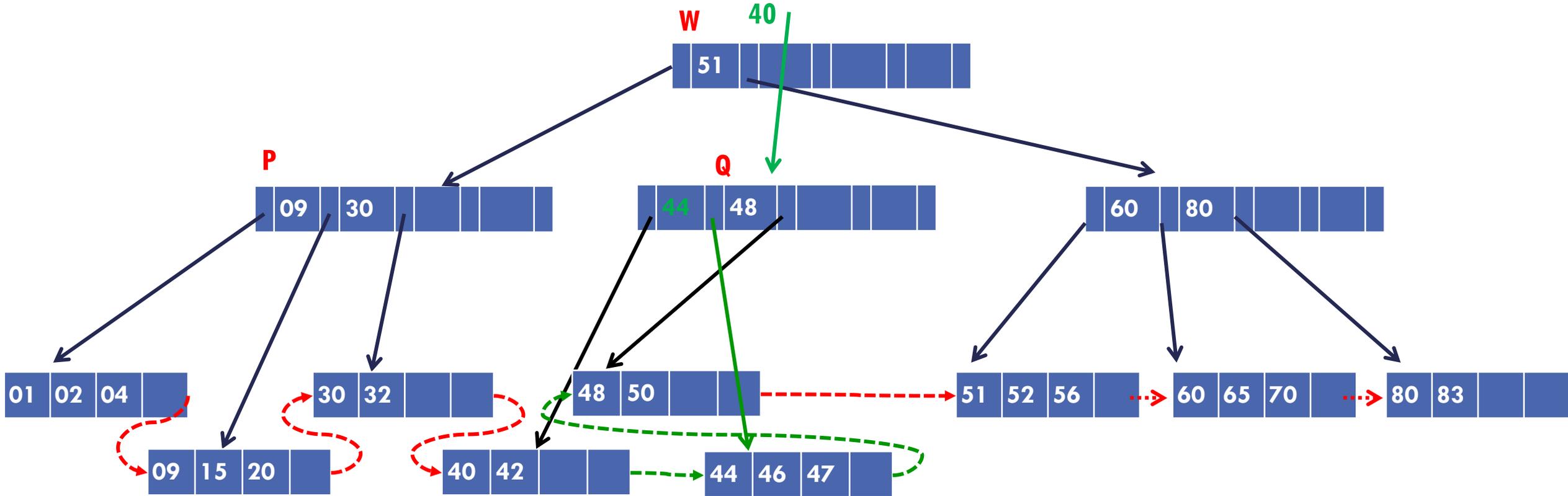
EXEMPLO DE INSERÇÃO COM PARTICIONAMENTO DE NÓ INTERNO: INSERIR CHAVE 44

Criar nova página Q



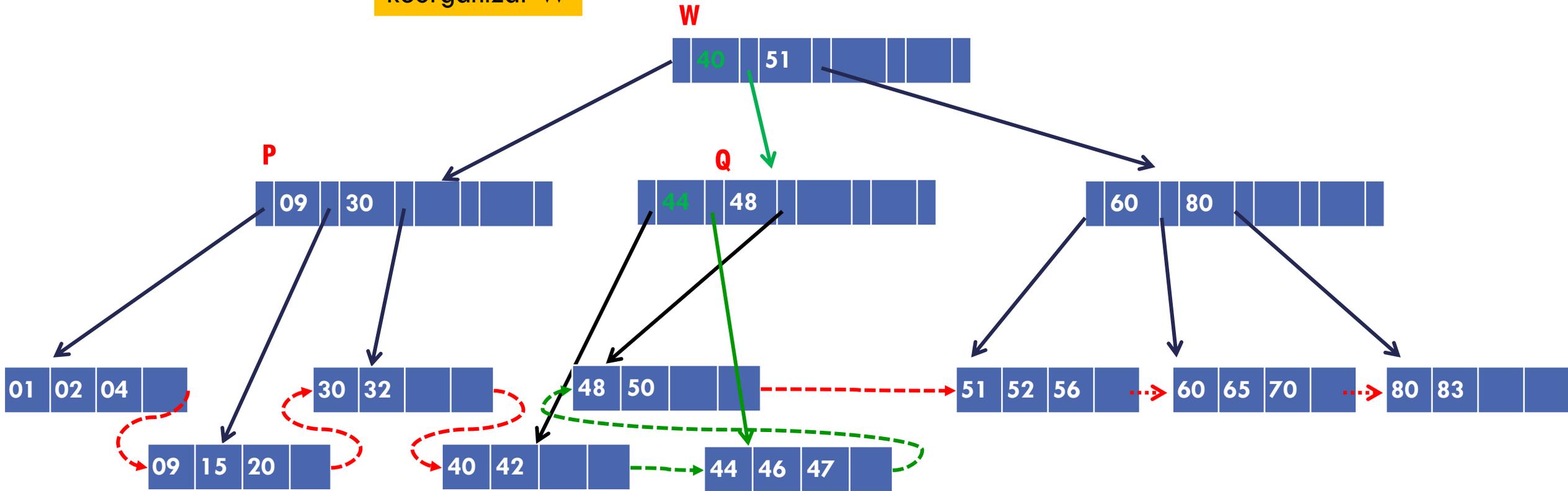
Dividir as chaves entre as duas páginas (09; 30; 40; 44; 48)
 d chaves na página original P
 chave d+1 sobe para nó pai W
 chaves d+2 em diante na nova página Q

COM PARTICIONAMENTO DE NÓ INTERNO: INSERIR CHAVE 44



EXEMPLO DE INSERÇÃO COM PARTICIONAMENTO DE NÓ INTERNO: INSERIR CHAVE 44

Reorganizar W



EXCLUSÃO

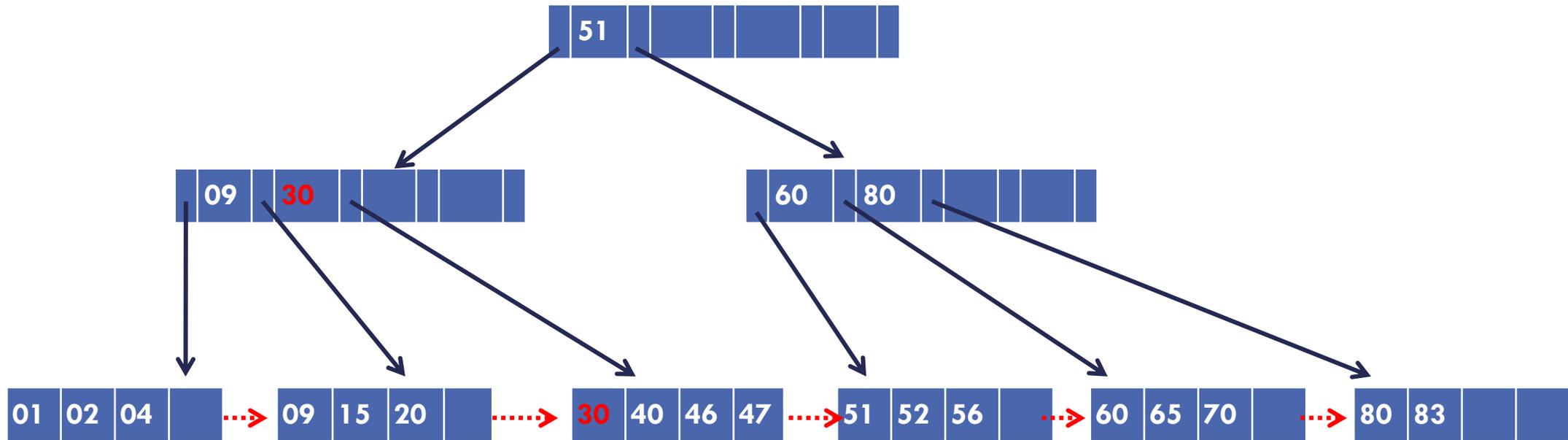
Excluir apenas no nó folha

Chaves excluídas continuam nos nós intermediários

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 30

ordem $d = 2$

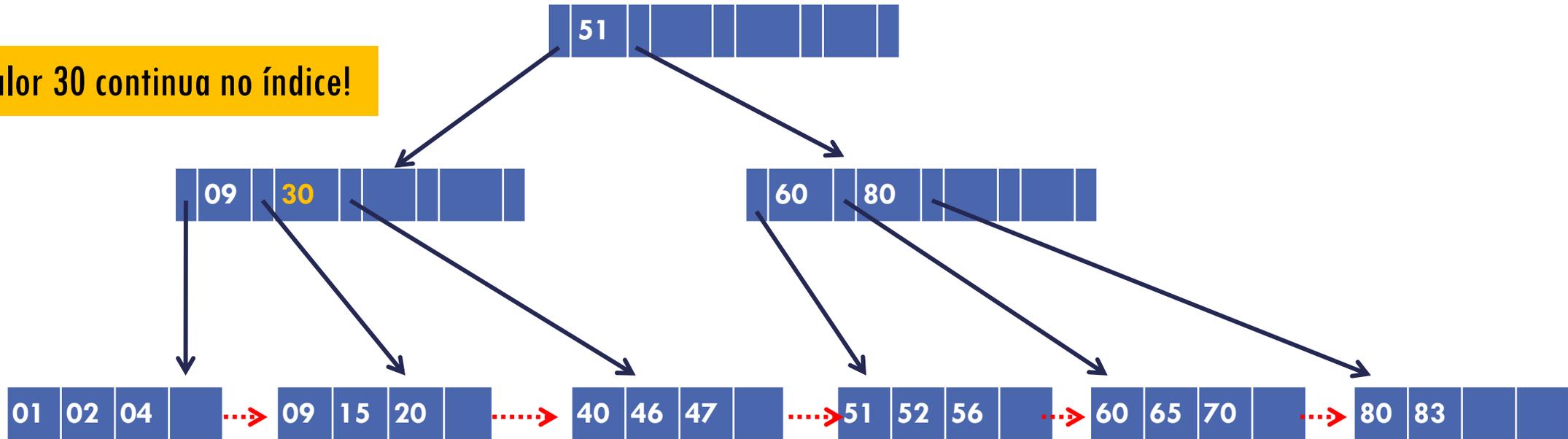


EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 30

ordem $d = 2$

O valor 30 continua no índice!



EXCLUSÃO QUE CAUSA CONCATENAÇÃO

Exclusões que causem concatenação de folhas podem se propagar para os nós internos da árvore

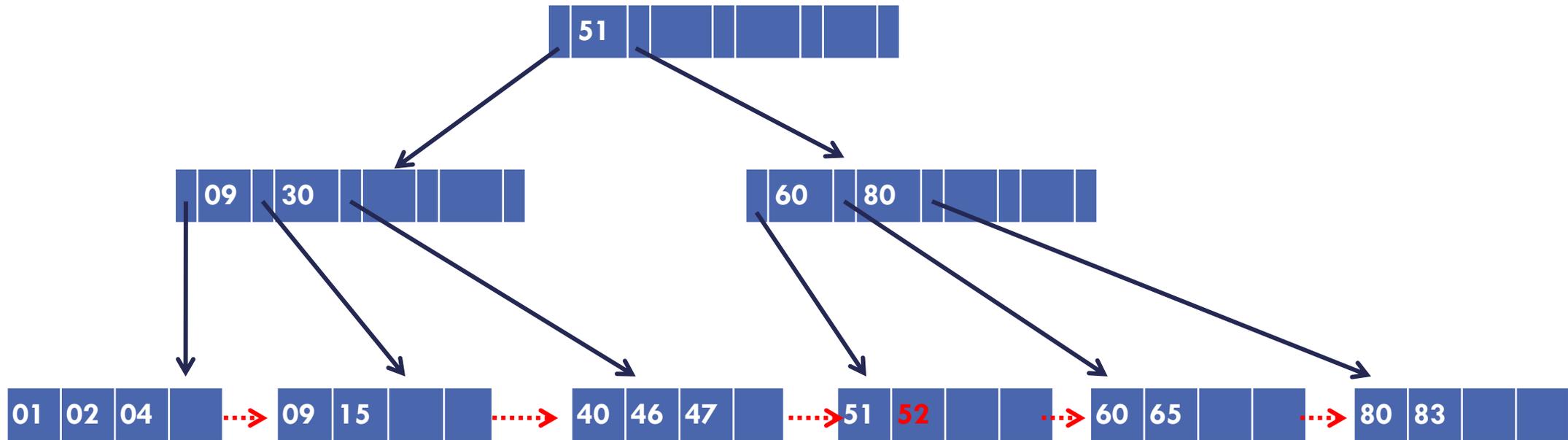
Importante:

- Se a concatenação ocorrer **na folha**: **a chave do nó pai não desce para o nó concatenado**, pois ele não carrega dados com ele. Ele é simplesmente apagado.
- Se a concatenação ocorrer em **nó interno**: usa-se a mesma lógica utilizada na **árvore B**

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

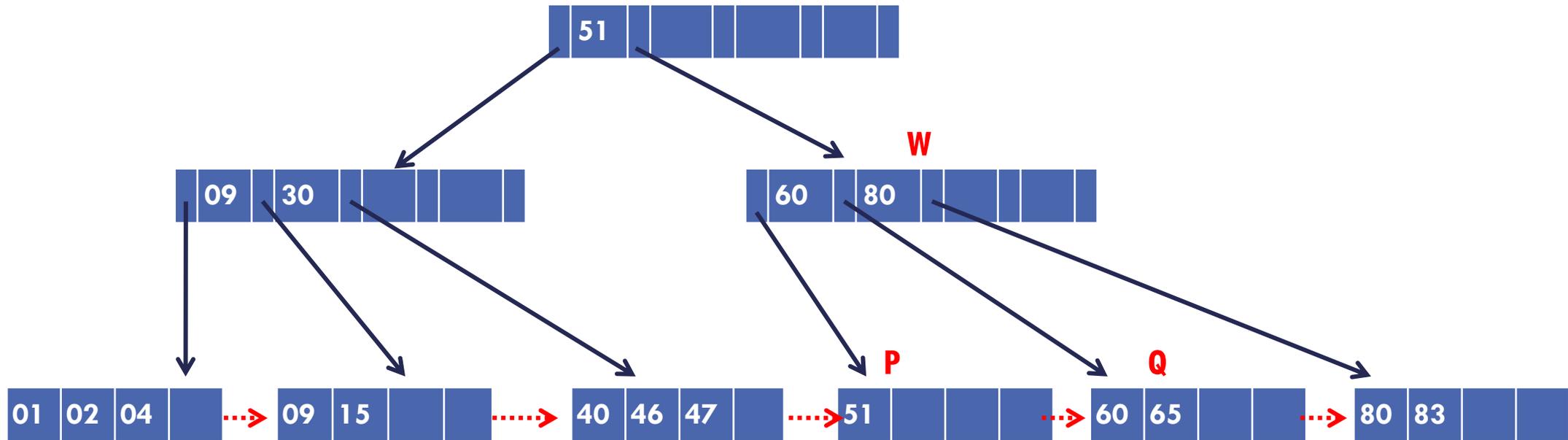
ordem $d = 2$



EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

ordem $d = 2$

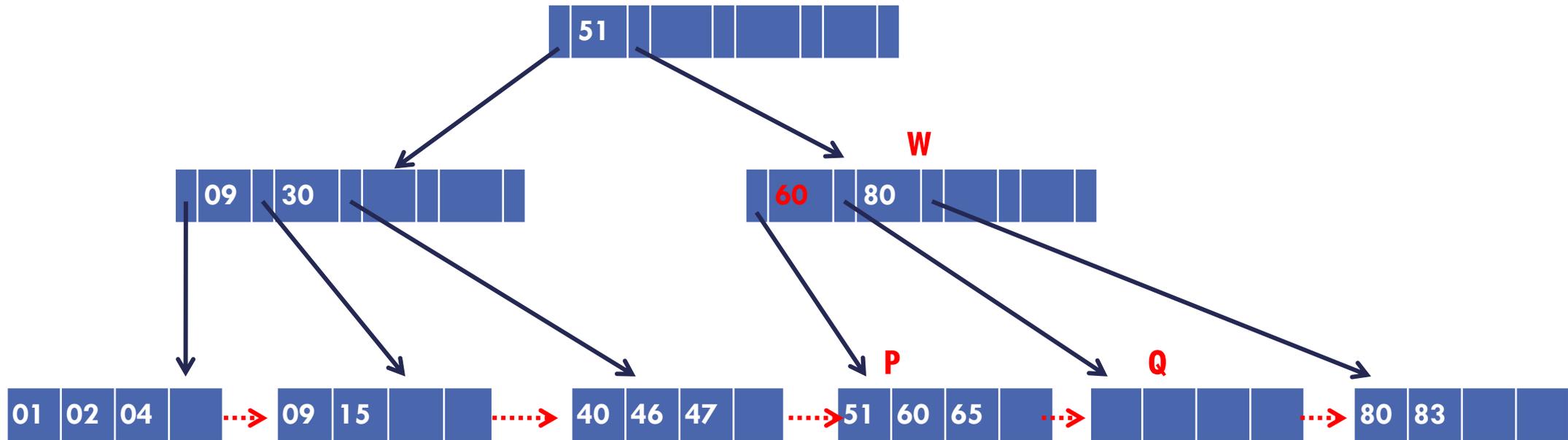


Nó ficou com menos de d entradas – necessário tratar isso
Soma dos registros de P e Q $< 2d$
Usar concatenação

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

ordem $d = 2$



Concatenação:

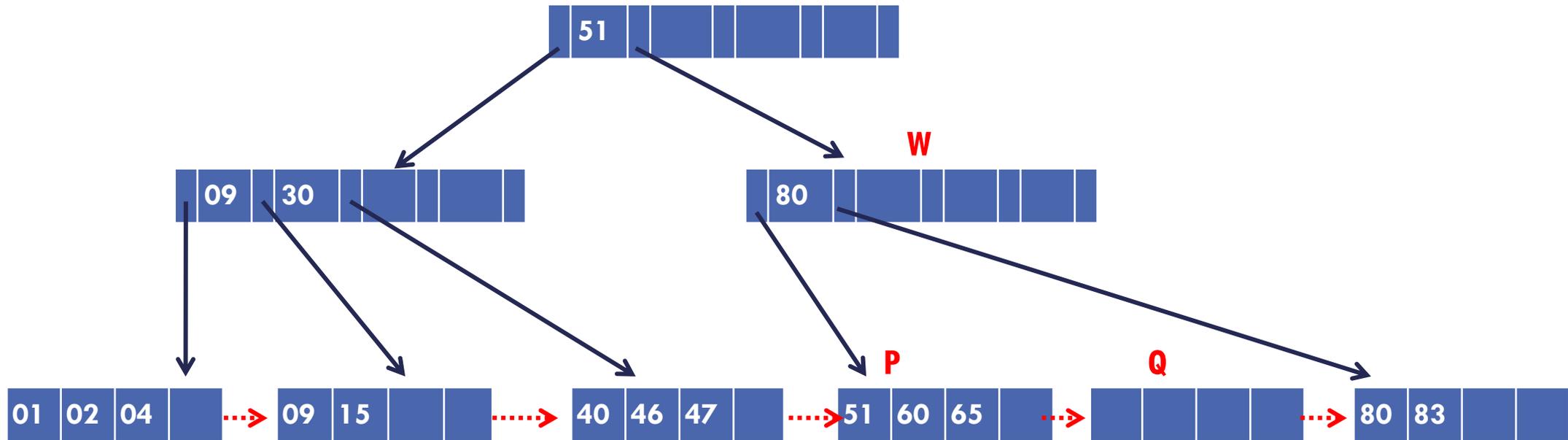
Passar os registros de Q para P

Eliminar a chave em W que divide os ponteiros para as páginas P e Q

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

ordem $d = 2$



Concatenação:

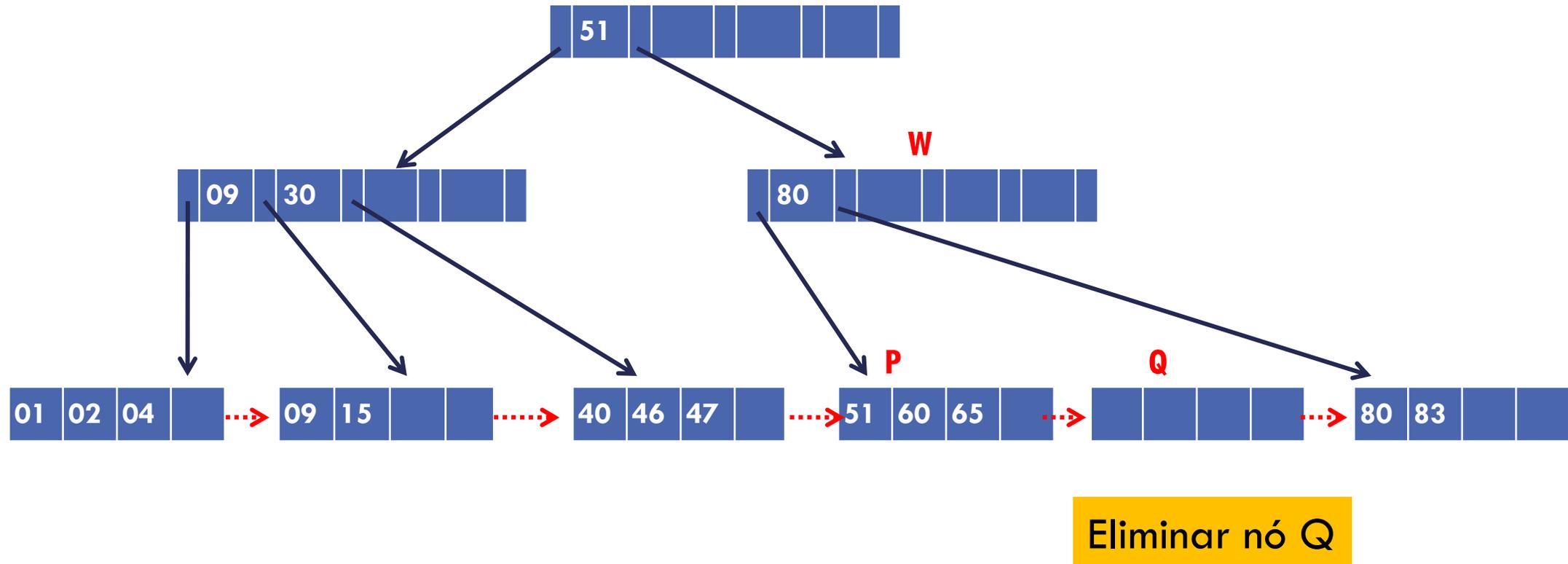
Passar os registros de Q para P

Eliminar a chave em W que divide os ponteiros para as páginas P e Q

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

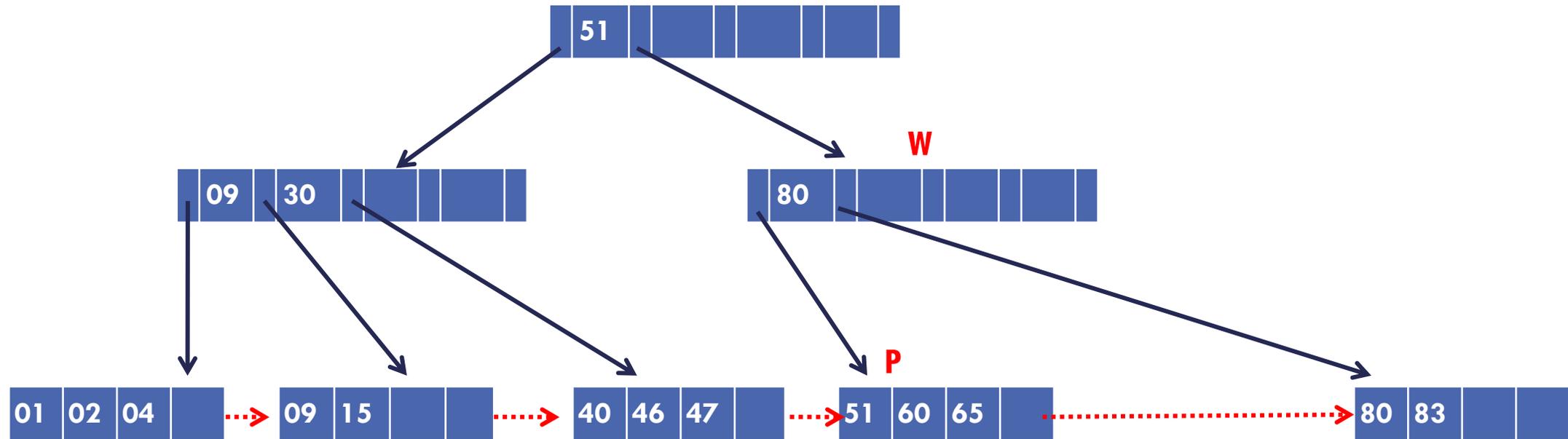
ordem $d = 2$



EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

ordem $d = 2$

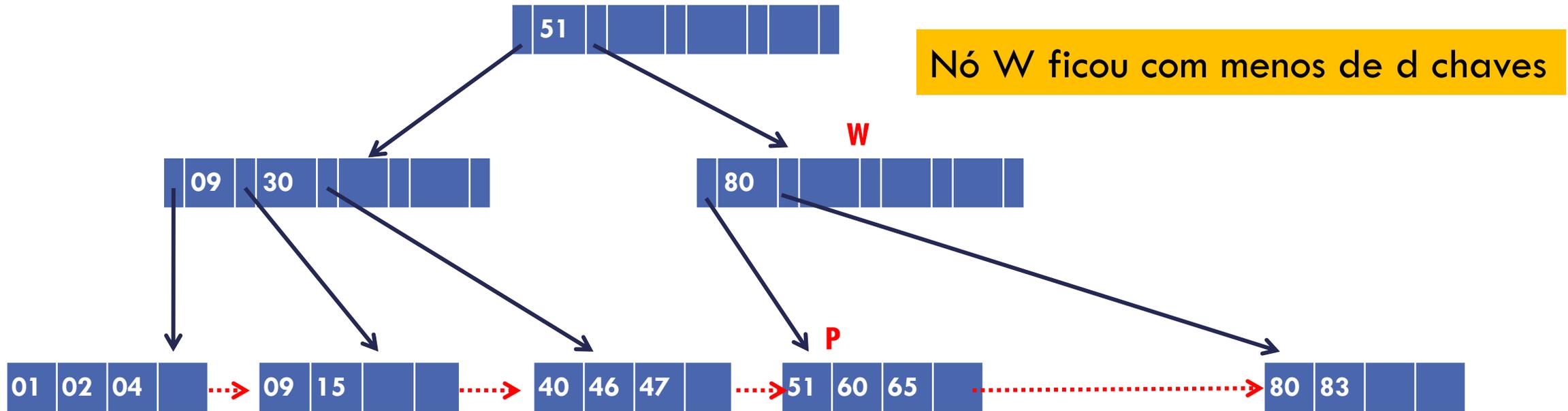


Eliminar nó Q

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

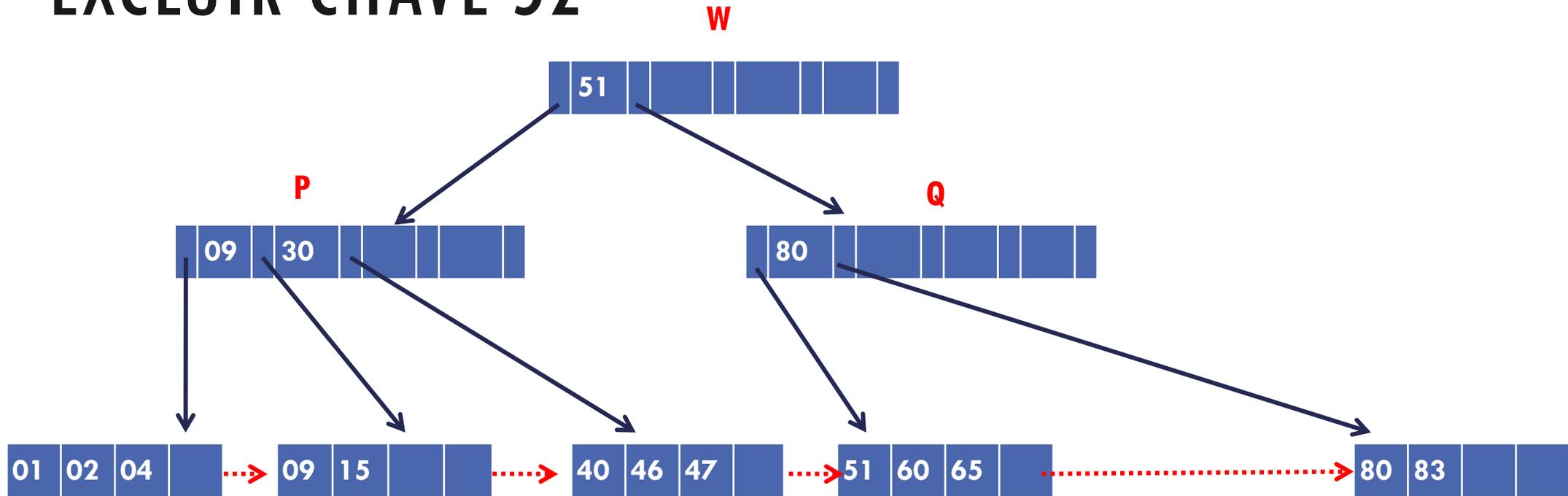
ordem $d = 2$



EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

ordem $d = 2$

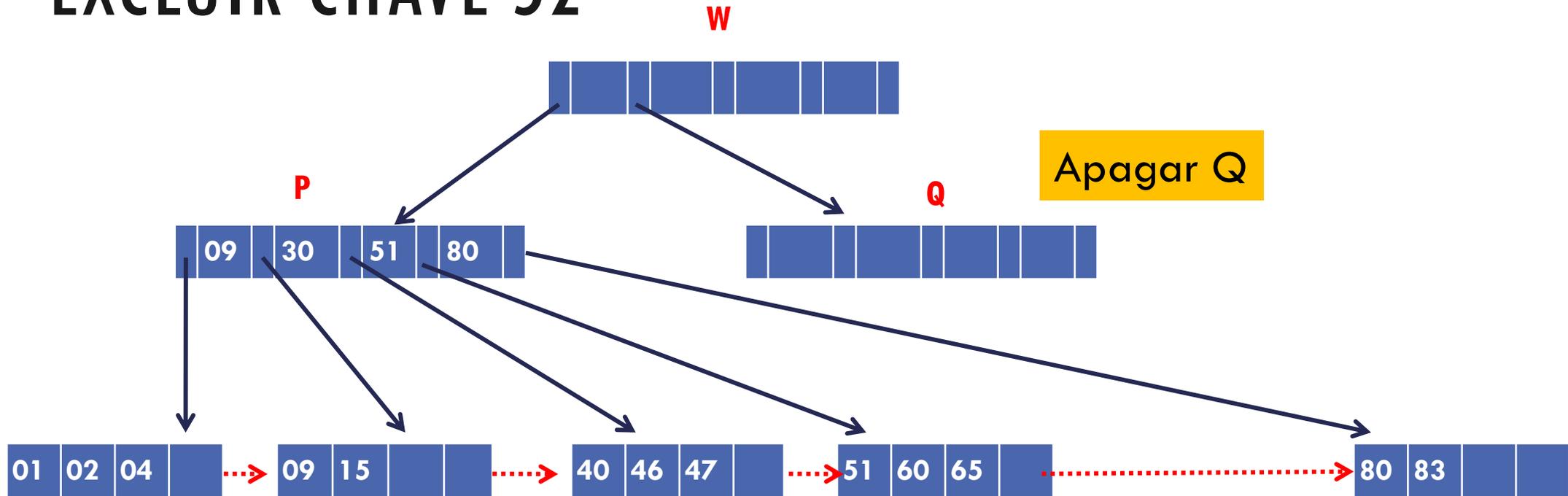


Soma de total de chaves de P e Q $< 2d$
Solução: concatenação

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

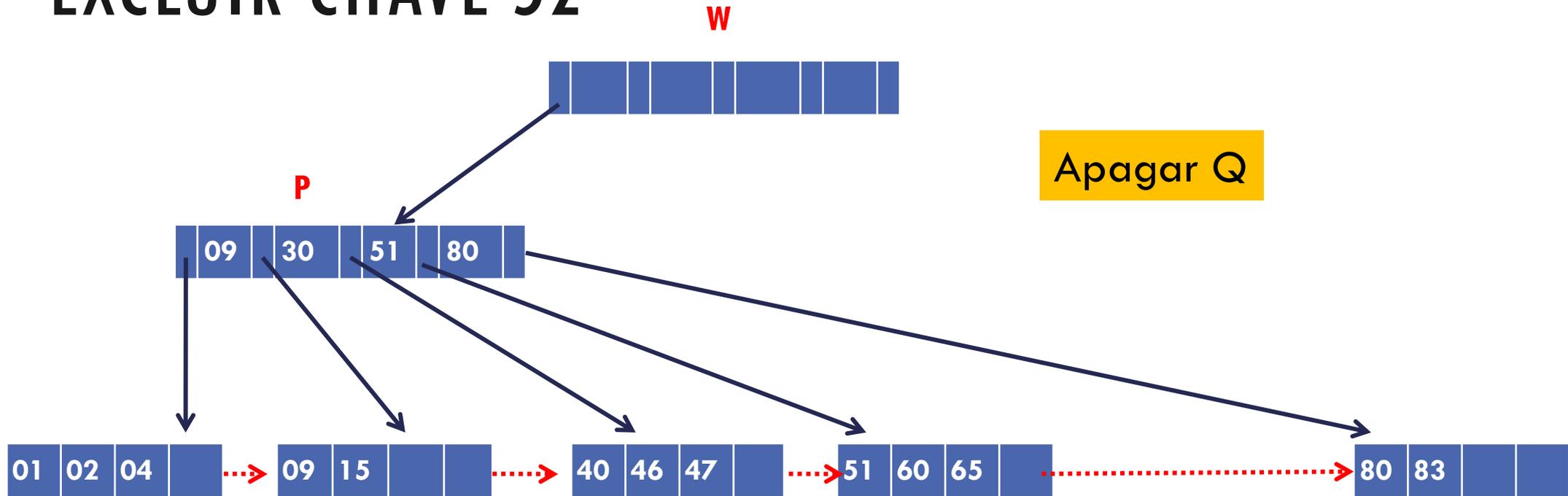
ordem $d = 2$



EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

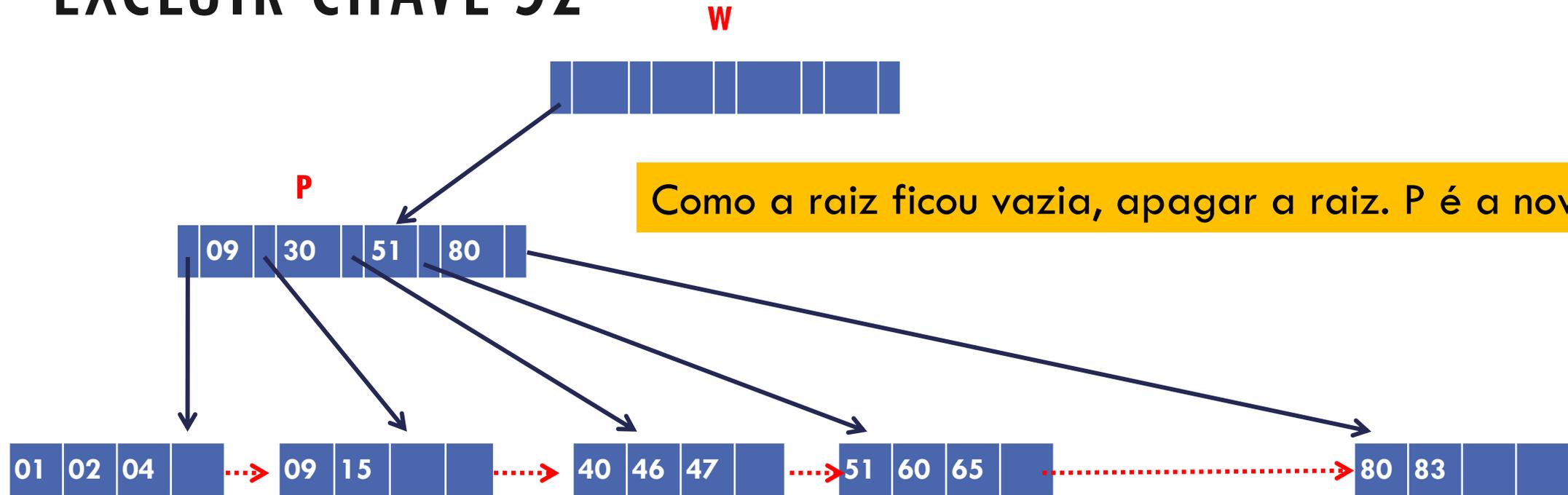
ordem $d = 2$



EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

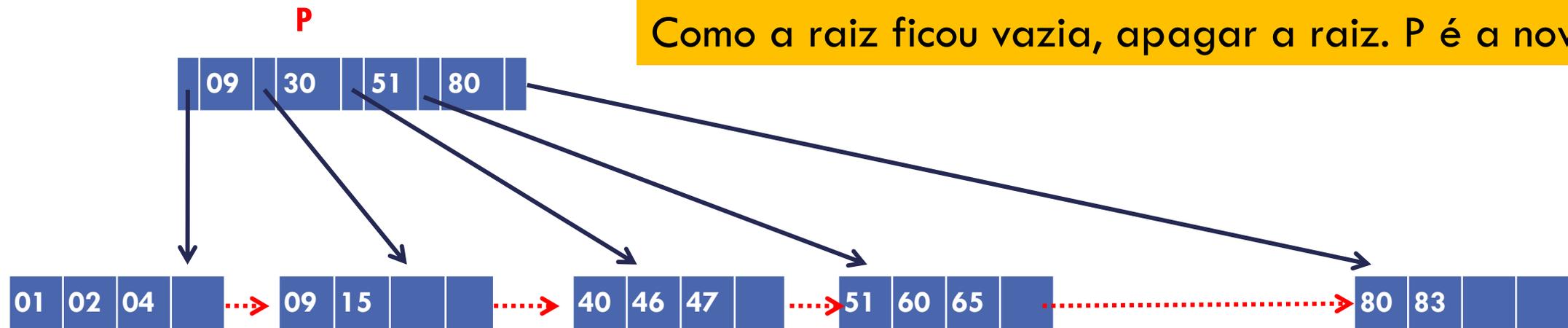
ordem $d = 2$



EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 52

ordem $d = 2$



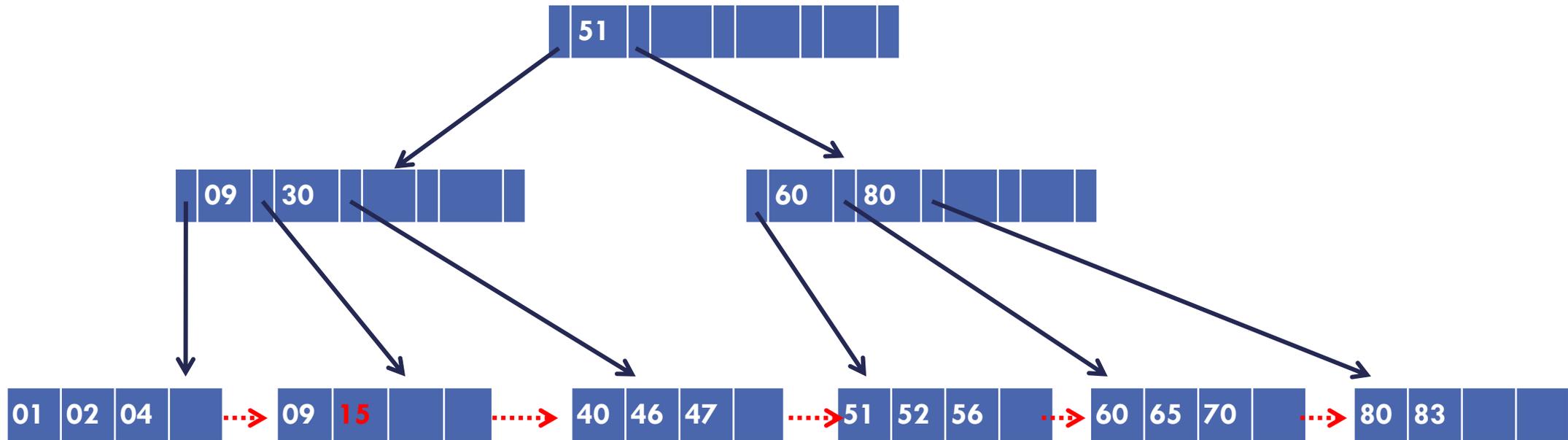
EXCLUSÃO QUE CAUSA REDISTRIBUIÇÃO

Exclusões que causem redistribuição dos registros nas folhas provocam mudanças no conteúdo do índice, mas não na estrutura (não se propagam)

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 15

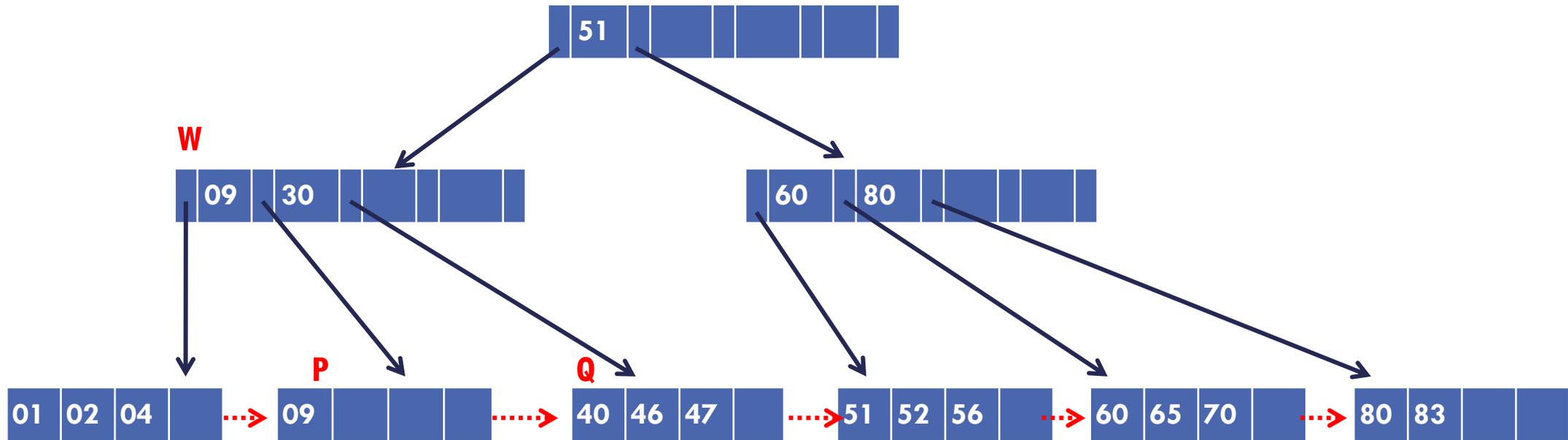
ordem $d = 2$



EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 15

ordem $d = 2$

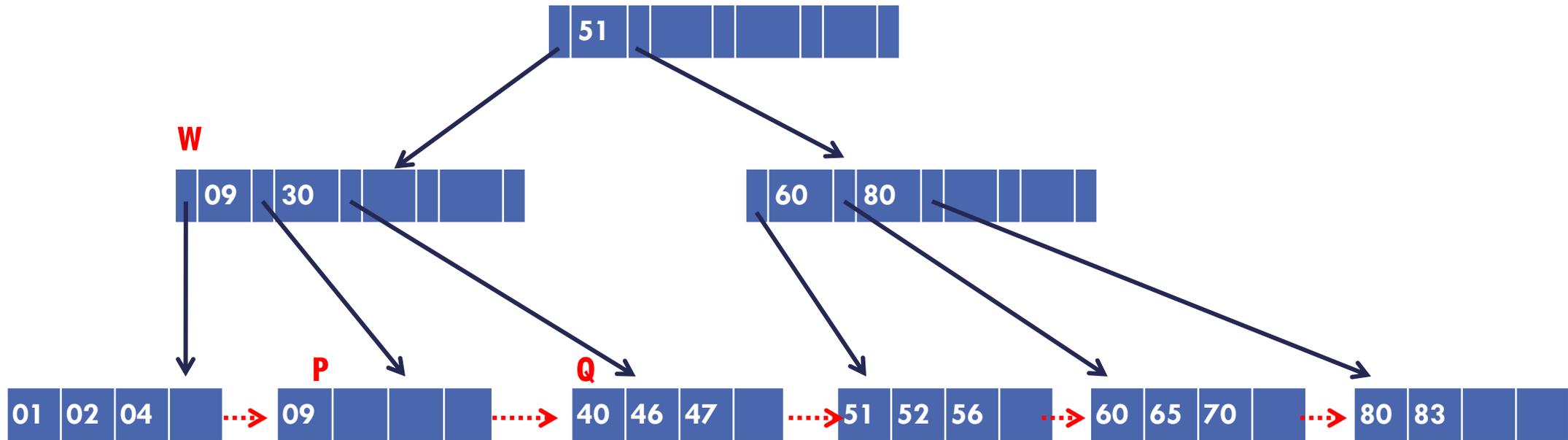


Nó ficou com menos de d entradas – necessário tratar isso
A soma dos registros dos irmãos adjacentes é $\geq 2d$
Solução: **redistribuição**
Como existem 2 opções, vamos optar pelo **nó da direita**

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 15

ordem $d = 2$

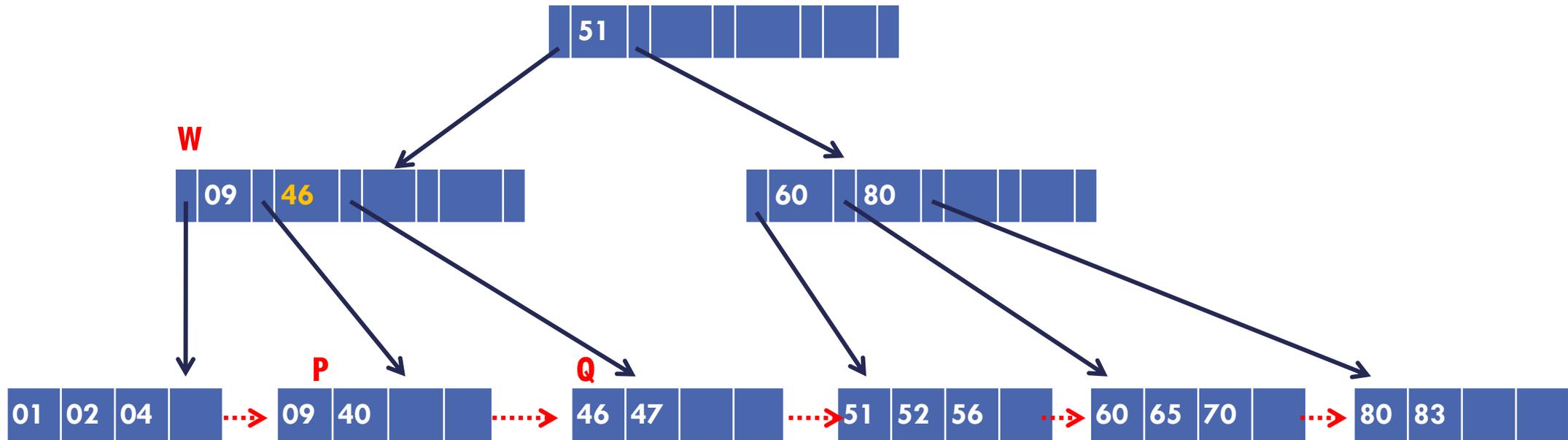


MAS... Se a chave do nó pai não precisa descer (porque não tem conteúdo, tem apenas a chave), porque não podemos concatenar P e Q, já que nesse exemplo a soma é $=2d$?
Resposta: ao concatenar P e Q, a página concatenada ficaria cheia, e a próxima inserção neste nó causaria um particionamento. Para evitar isso, continuamos obedecendo o critério: fazer concatenação apenas quando a soma da quantidade de chaves $< 2d$, e, sempre que tivermos as duas opções, optaremos pela redistribuição, que não se propaga.

EXEMPLO DE EXCLUSÃO EM ÁRVORE B+

EXCLUIR CHAVE 15

ordem $d = 2$



(09; 40; 46; 47)

d primeiras chaves ficam em P

Chave $d+1$ sobe para substituir a chave que já existia lá

Registros $d+1$ em diante ficam em Q

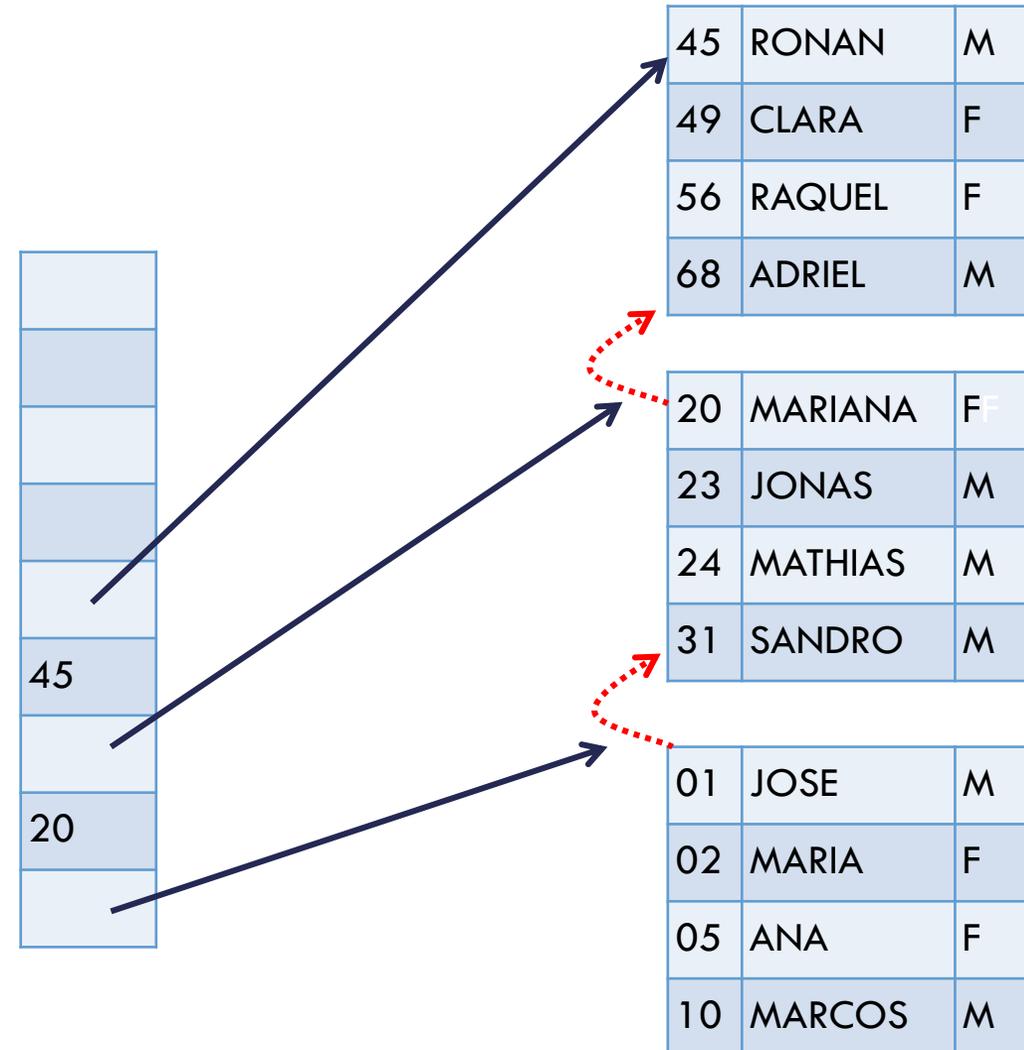
Note que a chave 46 sobe para W para substituir a chave 30, mas o registro correspondente é colocado em Q.

EXEMPLO

(MOSTRANDO OS DADOS NAS FOLHAS)

Neste exemplo, a árvore B+ tem o nó raiz e 3 folhas

Ordem da árvore é $d = 2$



CONSIDERAÇÕES SOBRE IMPLEMENTAÇÃO EM DISCO

Pode-se utilizar três arquivos:

- Um arquivo para armazenar os metadados
 - Ponteiro para a raiz da árvore
 - Flag indicando se a raiz é folha
- Um arquivo para armazenar o índice (nós internos da árvore)
- Um arquivo para armazenar os dados (folhas da árvore)

ESTRUTURA DO ARQUIVO DE ÍNDICE

O arquivo de índice estará estruturado em nós (blocos/páginas)

Ele contém os nós **internos** da árvore

Cada nó possui

- Inteiro representando o número de chaves (**m**) armazenadas no nó
- **Flag** booleano que diz se página aponta para nó folha (**TRUE** se sim, **FALSE** se não)
- Ponteiro para o nó pai (para facilitar a implementação de concatenação)
- $p_0, (s_1, p_1), (s_2, p_2), \dots, (s_d, p_d), (s_{d+1}, p_{d+1}), \dots, (s_{2d+1}, p_{2d+1})$, onde:
 - p_i é um ponteiro para uma página (dentro deste arquivo, se **flag** é **FALSE**, no arquivo de dados, se **flag** é **TRUE**)
 - s_i é uma chave

ESTRUTURA DO ARQUIVO DE DADOS

O arquivo de dados também estará estruturado em nós (blocos/páginas)

Ele contém os nós **folhas** da árvore

Cada nó possui

- Inteiro representando o número de registros (**m**) armazenadas no nó
- Ponteiro para o nó pai (para facilitar a implementação de concatenação)
- Ponteiro para a próxima página
- **m** registros

CONSIDERAÇÕES SOBRE IMPLEMENTAÇÃO

Se o sistema de armazenamento tem tamanho de bloco de **B** bytes, e as chaves a serem armazenadas têm tamanho **k** bytes, a árvore B+ mais eficiente é a de ordem **$d = (B / k) - 1$**

Exemplo prático:

- Tamanho do bloco do disco $B = 4\text{KB} = 4096$ bytes
- Tamanho da chave $k = 4$ bytes
- $d = (4096/4) - 1 = 1023$
- Quantas chaves cada nó da árvore comportará no máximo, nessa situação?
- $2d = 2046$ chaves!

EXERCÍCIO: ÁRVORE B+

- a) Desenhar uma árvore B+ de **ordem 2** que contenha registros com as seguintes chaves: 1, 2, 3, 8, 15, 35, 36, 38, 39, 41, 43, 45, 51, 59
- b) Sobre o resultado do exercício (a), excluir os registros de chave: 3, 38, 1, 41
- c) Sobre o resultado do exercício (b), incluir os registros de chave: 5, 14, 52, 53, 54

REFERÊNCIA

Szwarcfiter, J.; Markezon, L. Estruturas de Dados e seus Algoritmos, 3a. ed.
LTC. Cap. 5

AGRADECIMENTOS

Exemplo cedido por Renata Galante