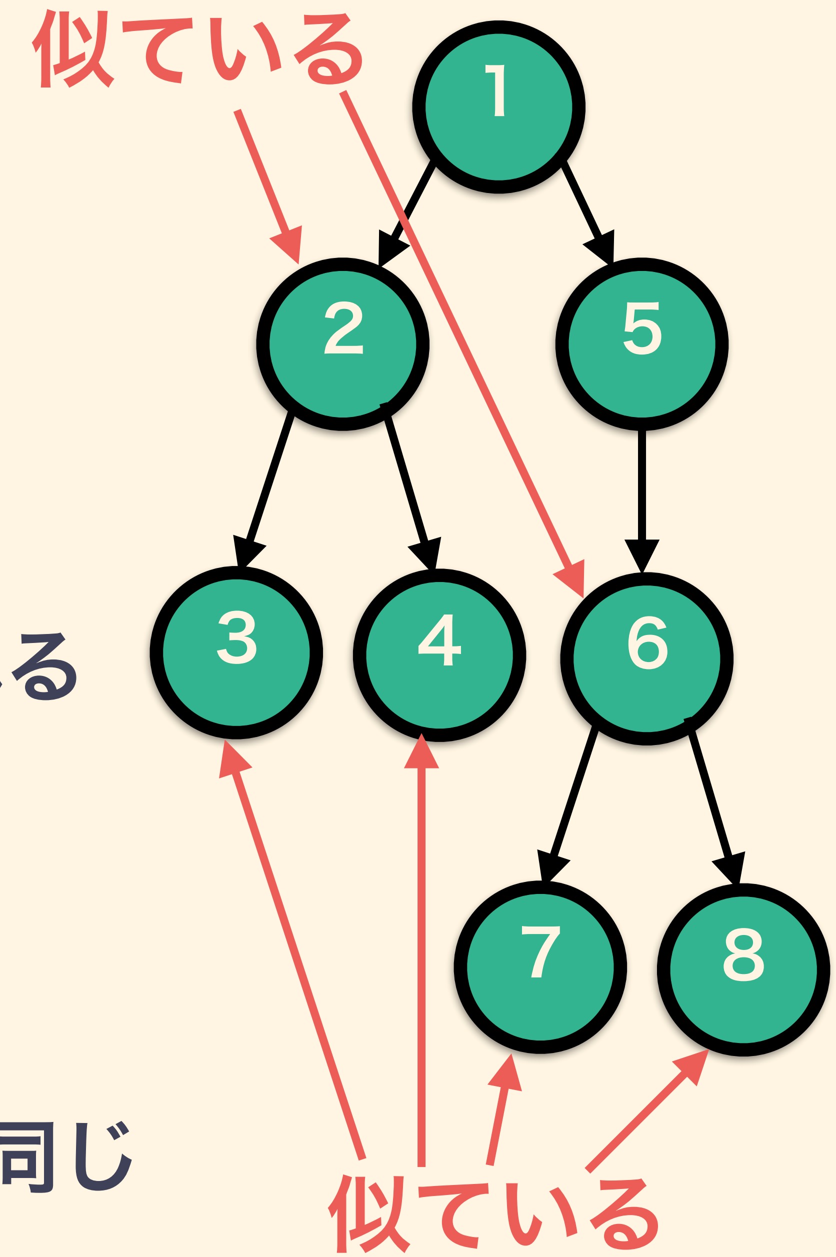
模擬地区予選2016
E: Similarity of Subtrees

原案：神谷
問題文：大橋
解答：井上・大橋・水野
解説：井上

問題概要



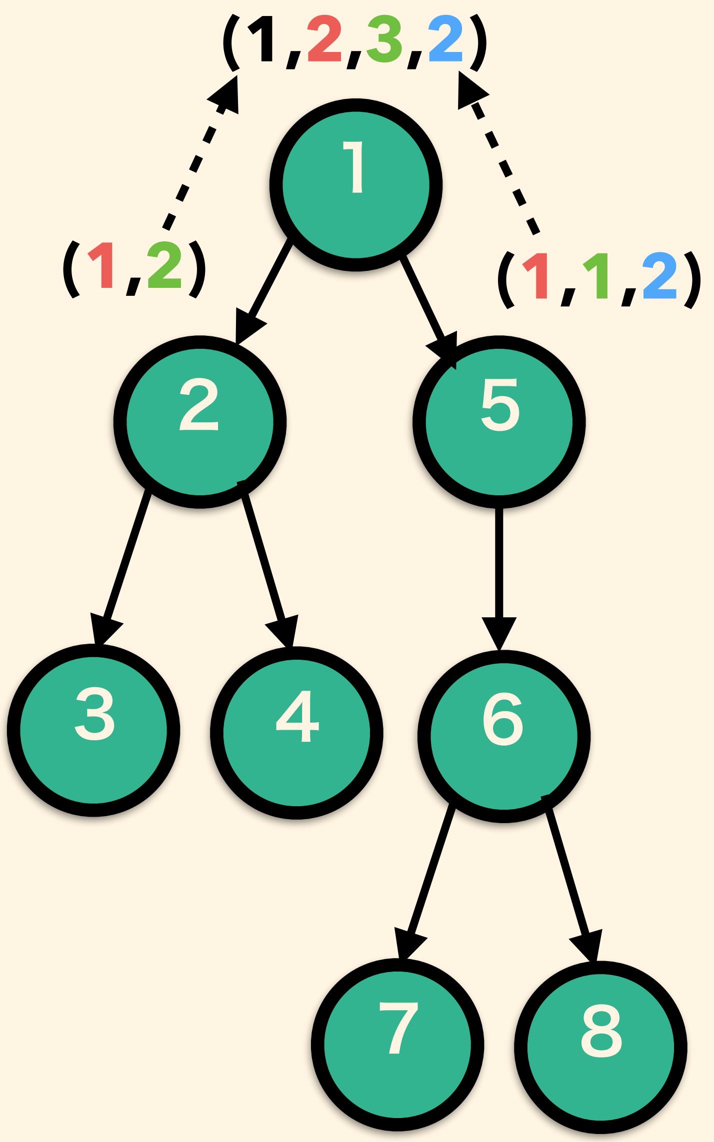
- N頂点の根付きラベル付き木 T が与えられる
  - ラベルは1~Nで重複なし、根は1
- 2 つの部分木が"似ている"

  ⇔ 各部分木が持つ同じ深さの子孫の数が同じ
- T の頂点 a,b (a<b) で、aを根とする部分木とbを根とする部分木が似ているような {a,b} の組はいくつある？

- 制約: $1 \leq N \leq 100{,}000$

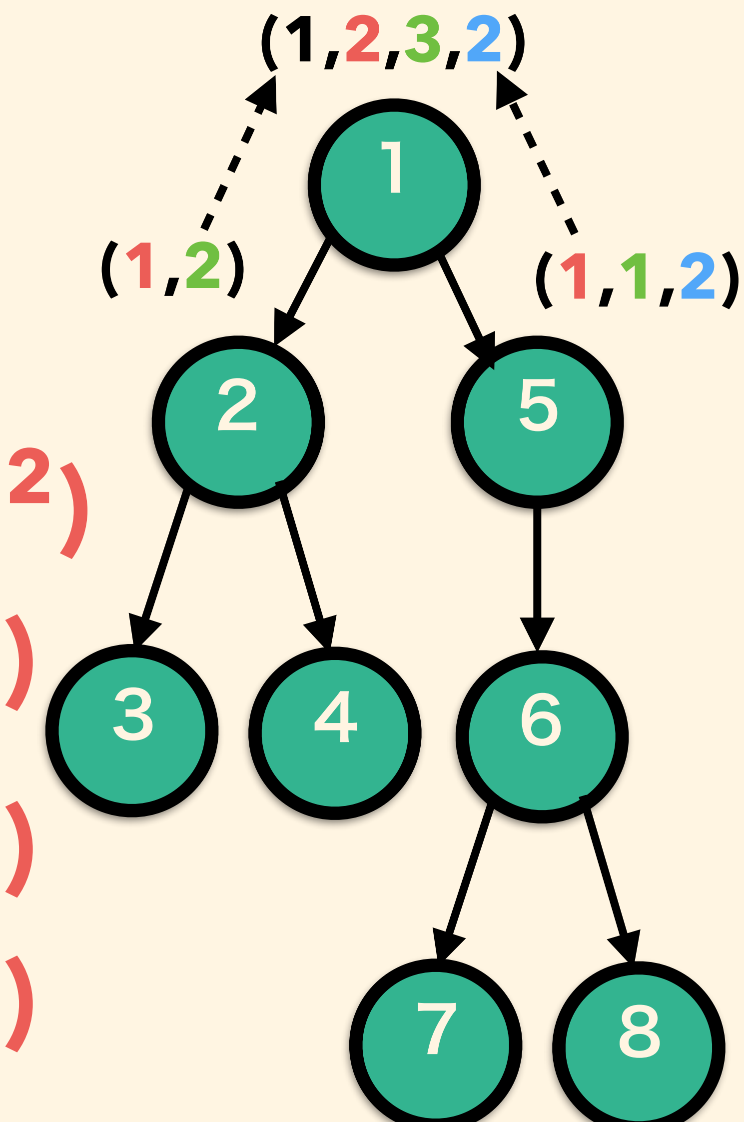
ナイーブ解法

- 各頂点ごとに、深さdの頂点がいくつあるかをd番目に記憶した配列を用意
- 類似判定: 各頂点ペアについて、
  配列が一致しているか判定
- 配列計算: 各頂点について、
  すべての子節点の配列を
  各深さごとに足しあわせ、
  先頭に1 (自分) を挿入



ナイーブ解法

- 各頂点ごとに、深さdの頂点がいくつあるかをd番目に記憶した配列を用意
- 類似判定: 各頂点ペアについて、 $O(N^2)$
  $O(N^3)$
  配列が一致しているか判定 $O(N)$
- 配列計算: すべての子節点の配列を $O(N)$
  $O(N^2)$
  各深さごとに足しあわせ、 $O(N)$
  先頭に1 (自分) を挿入 デックで $O(1)$



改善1: ハッシュ

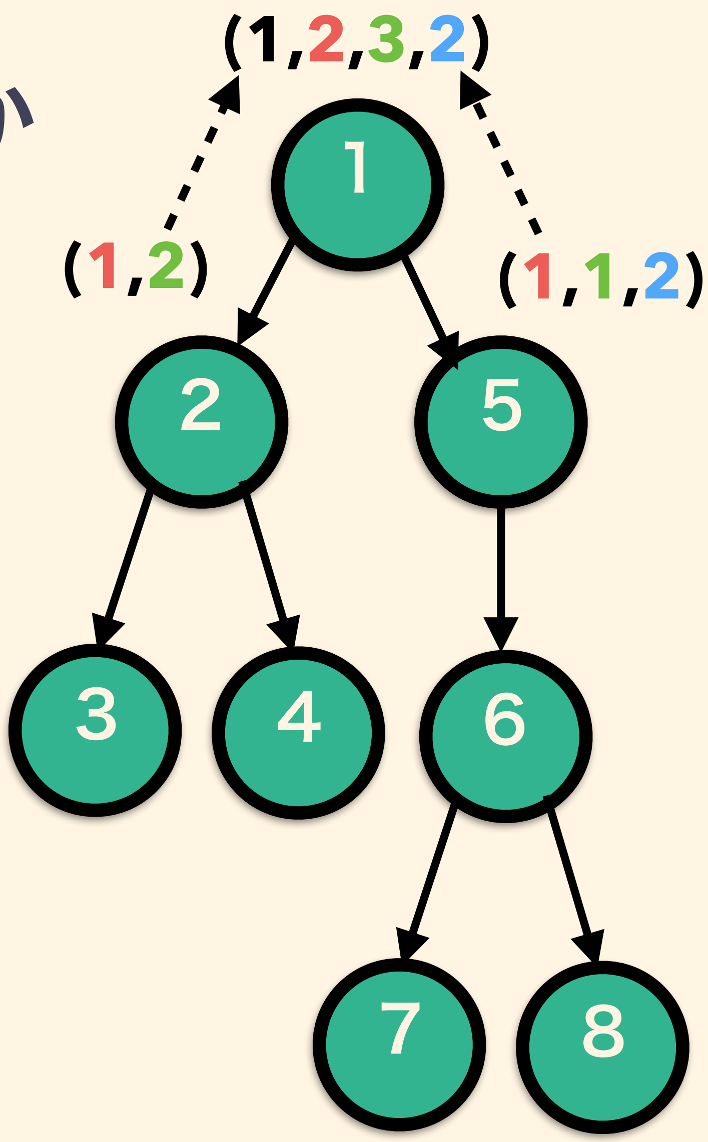
- 類似判定にハッシュを用いる
- 各配列に対応するハッシュ値を計算し、ハッシュにヒットしたらペアが作れる
- あるハッシュ値 h が出た数 x を覚えておけば、類似するペアが x(x-1)/2 出たことがわかる

(1,2) (1,2) (1,3,2)

ハッシュ値	0	1	2	3	4	5	6	7	8	9
個数	0	0	2	0	0	0	1	0	0	0

ちょっと改善された解法

- 各頂点ごとに、深さdの頂点がいくつあるかをd番目に記憶した配列を用意
- 類似判定: 各頂点について、 **O(N)**
  O(N)
  ハッシュの記録を1足す **O(1)**
- 配列計算: すべての子節点の配列を **O(N)**
  O(N^2)
  各深さごとに足しあわせ、**O(N)**
  先頭に1 (自分) を挿入して **O(1)**
  できた配列のハッシュ値を計算 **O(N)**



改善2: ローリングハッシュ

- ハッシュ関数としてローリングハッシュを用いる
- 配列Aのローリングハッシュ h(A):

  ある数b (大抵素数), m (大抵mと素な大きい数)を用いて、

  $h(A) = (\Sigma\ A[k] \times b^k) \bmod m$
- このハッシュ関数を用いると、ハッシュ値の計算が

  $h(A) = (\Sigma\ A[k] \times b^k) \bmod m$

  $= (\Sigma_{子節点の配列C}\ (B \times h(C)\ ) + 1) \bmod m$

  とできるので O(子節点数) で計算可能 → 全体でO(N)

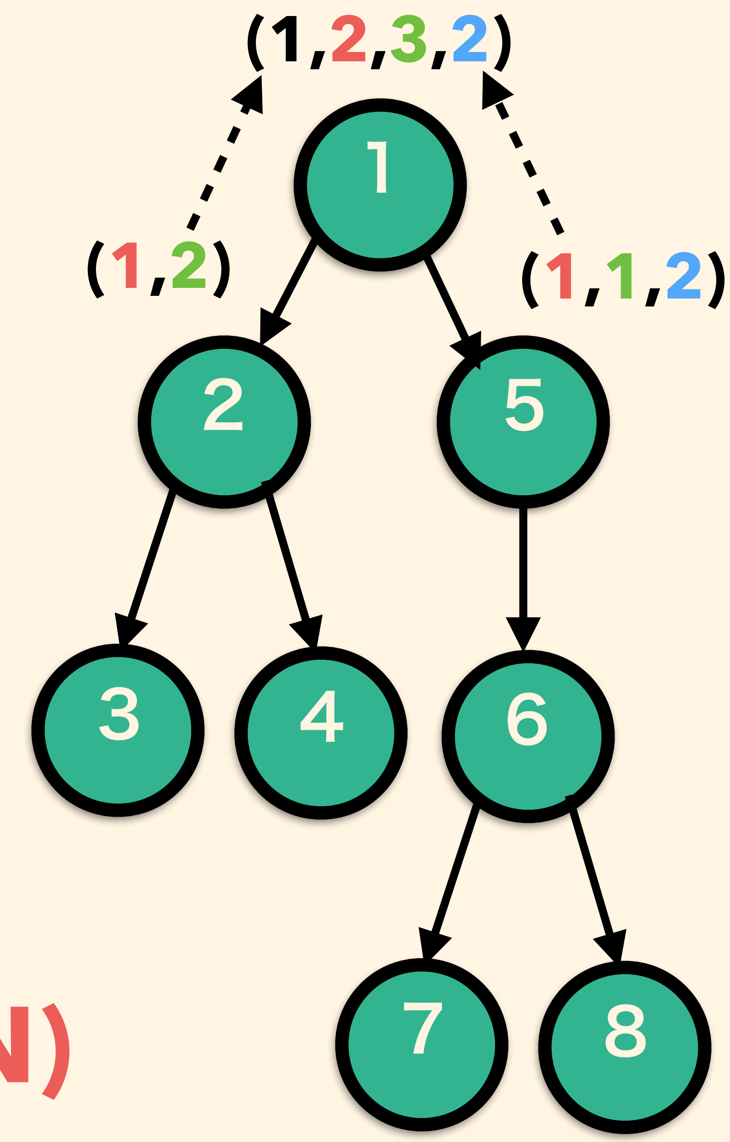
注意：ハッシュ値の衝突

- ハッシュ値が衝突する可能性があり、その場合は真面目に配列の比較をしないと厳密には同じ配列とはいえない
- mを十分大きくとることで、衝突する確率が十分低くなる 今回はこれを信用してOK
- たまによく使われるb, mについて衝突を狙うケースが用意されてる場合があるので注意
  - $b = 10^9+7$, $m = 2^{64}$ など
  - 複数のb, mを用いて多重チェックするなどで回避する

想定解法

- 各頂点ごとに、深さdの頂点がいくつあるかをd番目に記憶した配列を用意
- 類似判定: 各頂点について、 O(N)
  O(N)
  ハッシュの記録を1足す O(1)
- 配列計算: 子節点のハッシュ値から
  O(N)
  ハッシュ値を計算 全体でO(N)

(深さ配列を陽に持つ必要なし)



ジャッジ解

- 井上 (C++) : 41 行, 804 bytes
- 大橋 (python) : 37行, 821 bytes
- 水野 (C++) : 50行, 950 bytes

統計情報

- ACチーム数 / 提出チーム数
  - 23 / 79 (29.1%)
- First Acceptance
  - on-line: po (24:17)
  - on-site: po (24:17)