Design and Analysis of Algorithms Part IV: Graph Algorithms

Lecture 26: Strongly Connected Components

童咏昕

北京航空航天大学 计算机学院

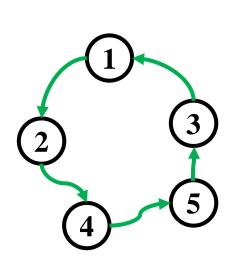
图算法篇概述



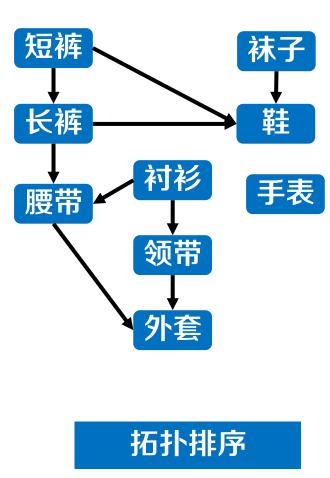
- 在算法课程第四部分"图算法"主题中,我们将主要聚焦于如下经典问题:
 - Basic Concepts in Graph Algorithms(图算法的基本概念)
 - Breadth-First Search (BFS, 广度优先搜索)
 - Depth-First Search (DFS, 深度优先搜索)
 - Cycle Detection (环路检测)
 - Topological Sort (拓扑排序)
 - Strongly Connected Components(强连通分量)
 - Minimum Spanning Trees (最小生成树)
 - Single Source Shortest Path (单源最短路径)
 - All-Pairs Shortest Paths (所有点对最短路径)
 - Bipartite Graph Matching (二分图匹配)
 - Maximum/Network Flows (最大流/网络流)

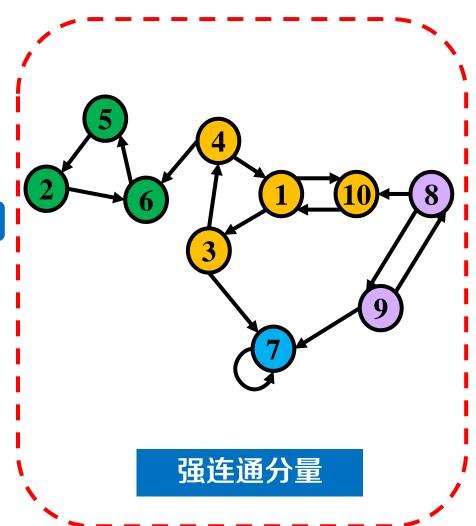
深度优先搜索应用





环路的存在性判断







问题背景与定义

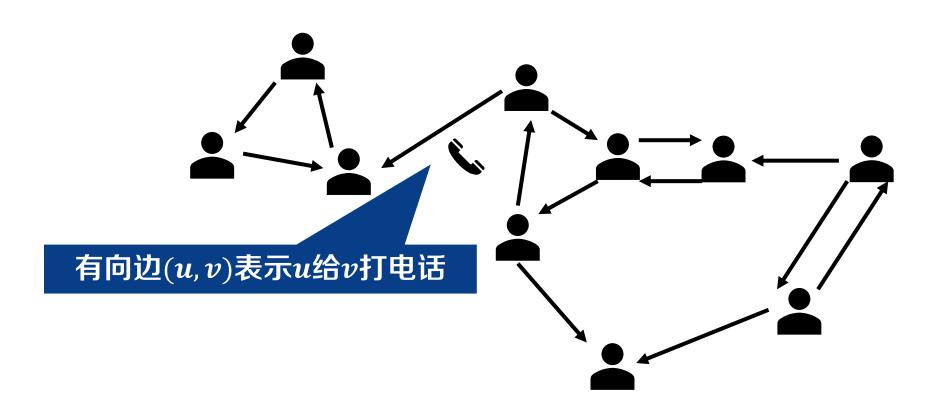
算法框架与实例

伪代码与复杂度

算法正确性证明

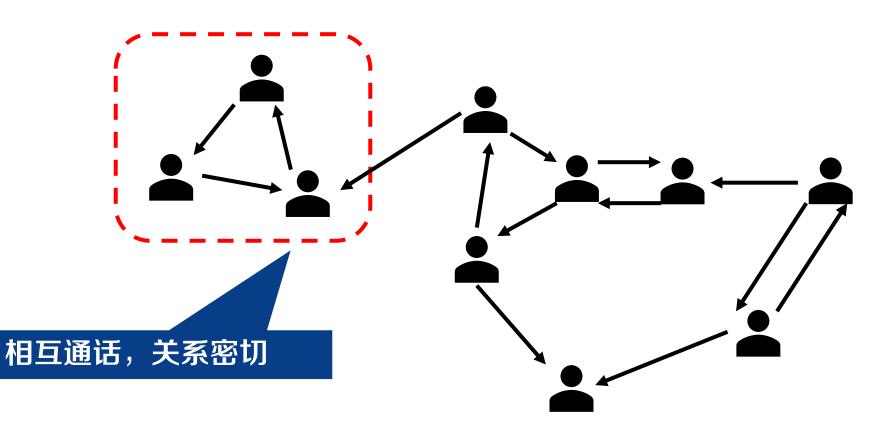


- 社交圏划分
 - 如何把人群按通话记录划分成不同的社交圈?



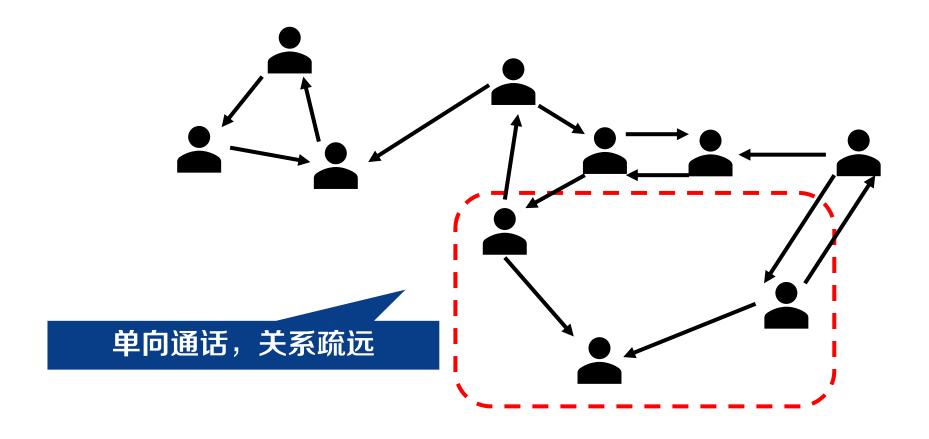


- 社交圏划分
 - 如何把人群按通话记录划分成不同的社交圈?



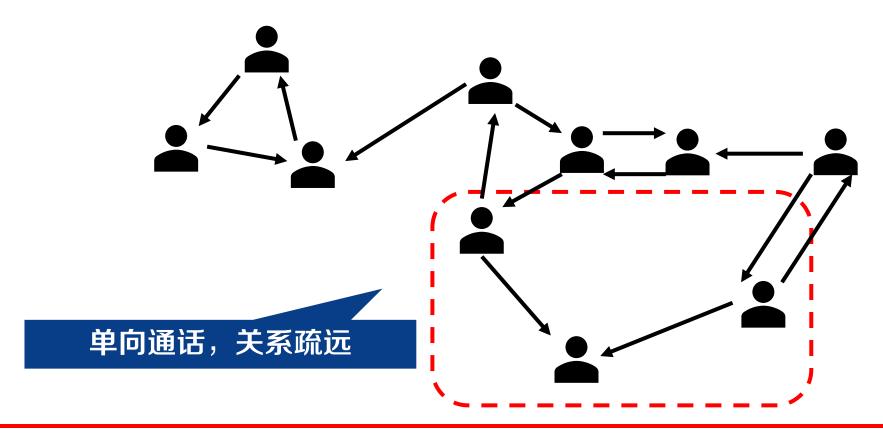


- 社交圏划分
 - 如何把人群按通话记录划分成不同的社交圈?





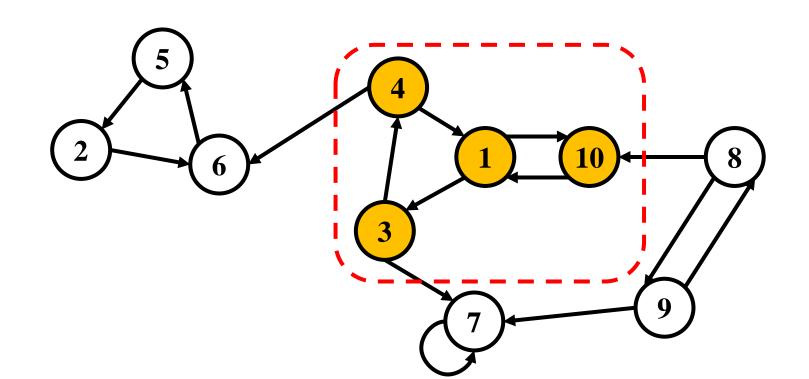
- 社交圏划分
 - 如何把人群按通话记录划分成不同的社交圈?



问题: 如何严格定义关系的亲密程度?

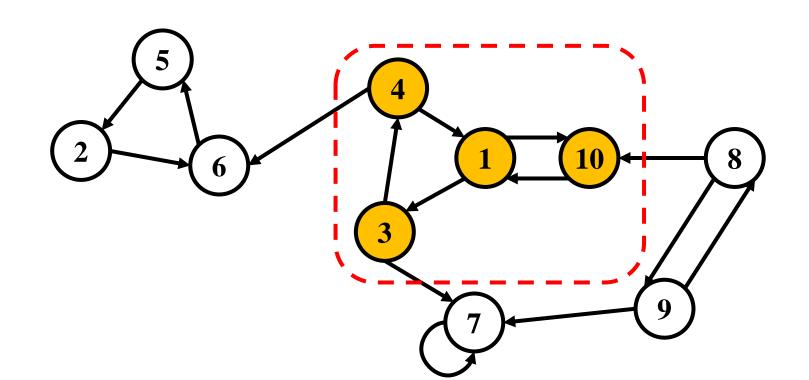


- 社交圏划分
 - 如何把人群按通话记录划分成不同的社交圈?
- 强连通分量
 - 一个强连通分量是顶点的子集



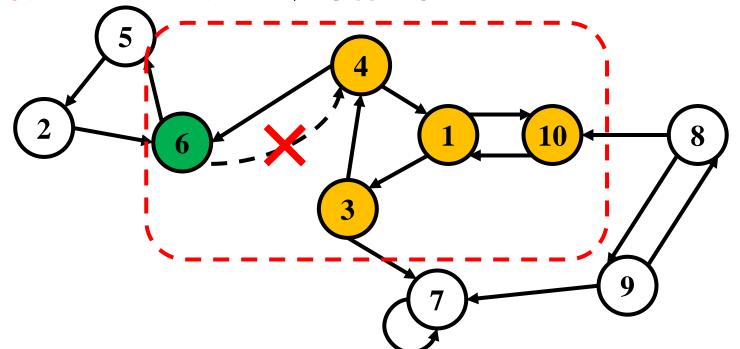


- 社交圏划分
 - 如何把人群按通话记录划分成不同的社交圈?
- 强连通分量
 - 一个强连通分量是顶点的子集
 - 强连通分量中任意两点相互可达



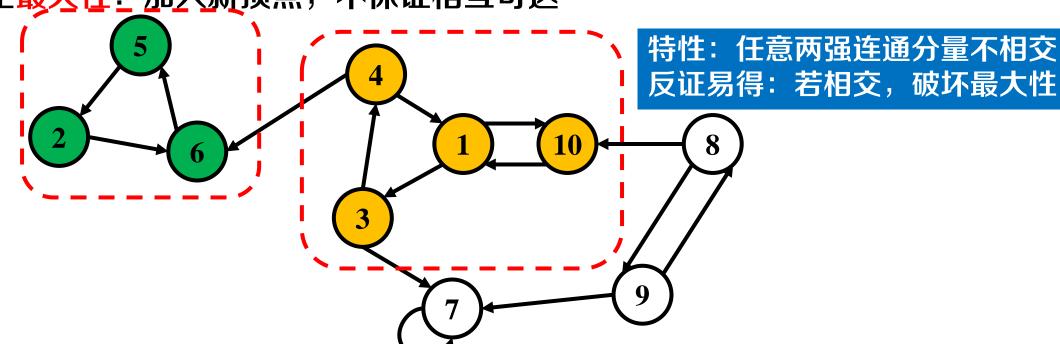


- 社交圈划分
 - 如何把人群按通话记录划分成不同的社交圈?
- 强连通分量
 - 一个强连通分量是顶点的子集
 - 强连通分量中任意两点相互可达
 - 满足最大性:加入新顶点,不保证相互可达





- 社交圏划分
 - 如何把人群按通话记录划分成不同的社交圈?
- 强连通分量
 - 一个强连通分量是顶点的子集
 - 强连通分量中任意两点相互可达
 - 满足最大性: 加入新顶点,不保证相互可达



问题定义



强连通分量

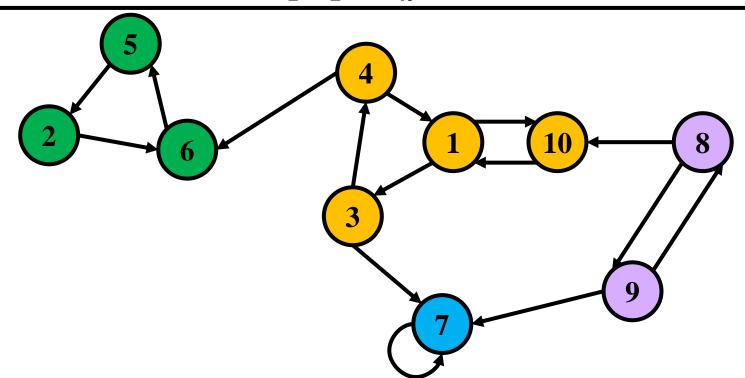
Strongly Connected Components

输人

• 有向图*G* =< *V*, *E* >

输出

• 图的所有强连通分量 $C_1, C_2, \dots C_n$





问题背景与定义

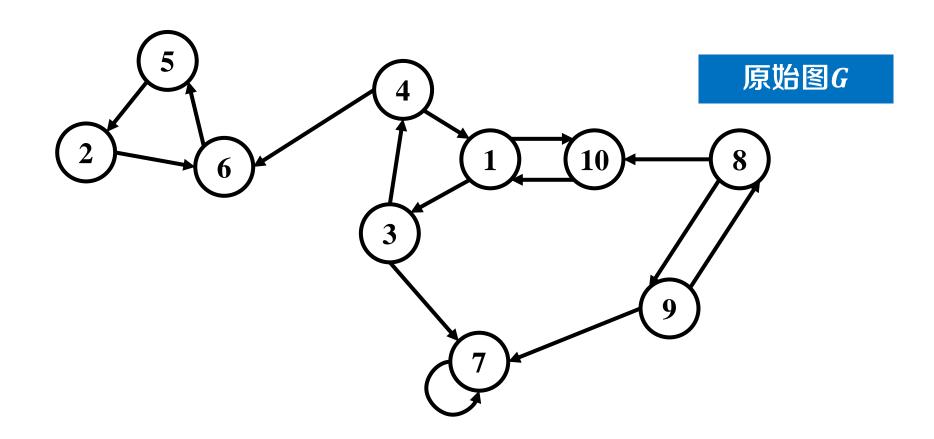
算法框架与实例

伪代码与复杂度

算法正确性证明

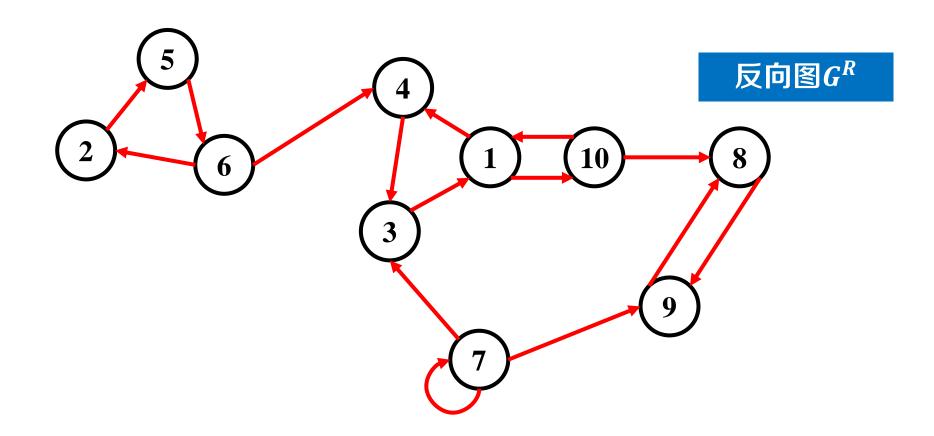


• 步骤1: 把边反向,得到反向图 G^R



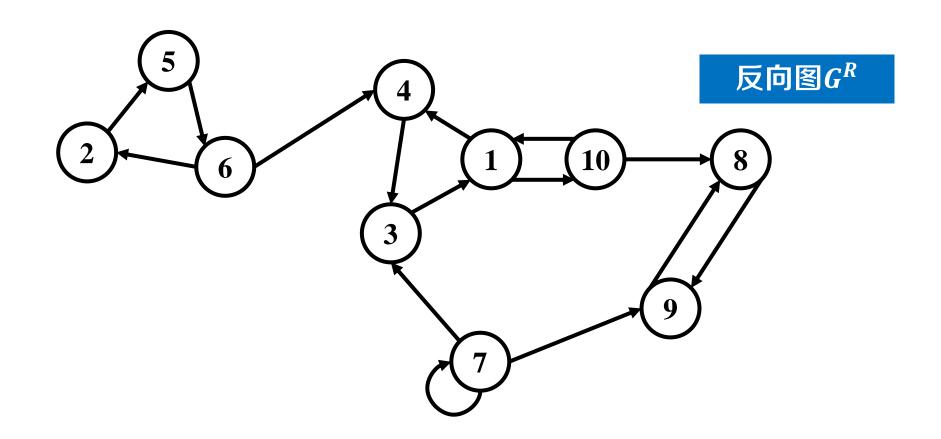


• 步骤1: 把边反向,得到反向图 G^R



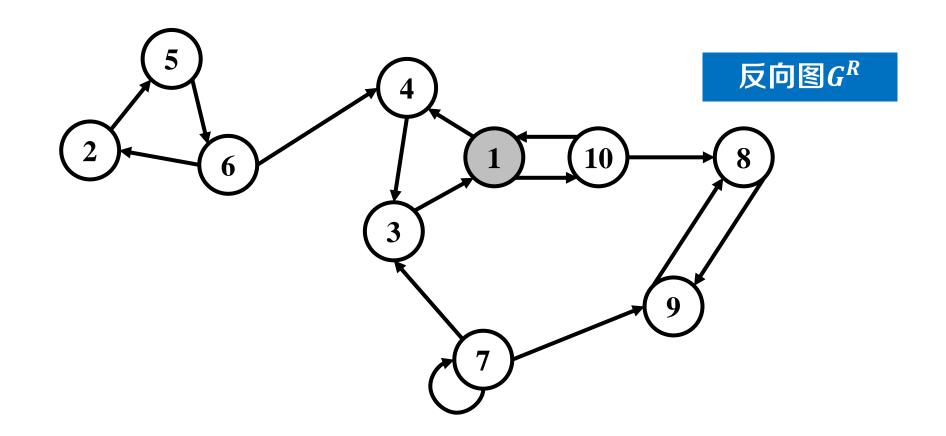


• 步骤1: 把边反向,得到反向图 G^R



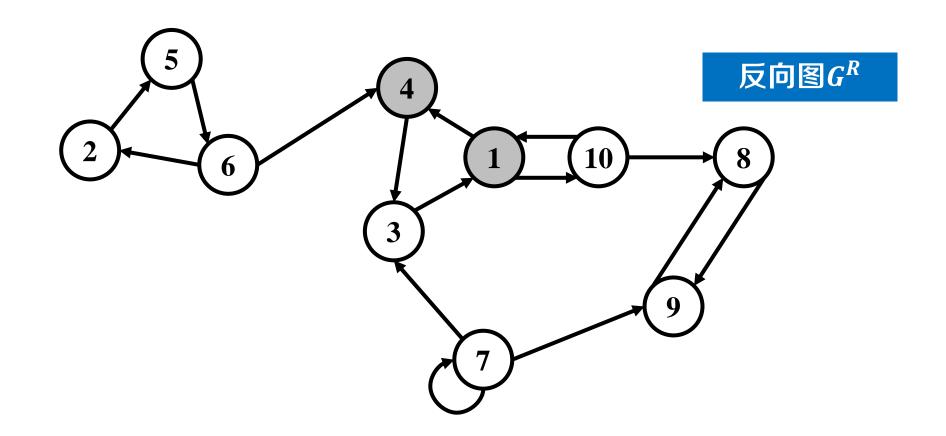


• 步骤1: 把边反向,得到反向图 G^R



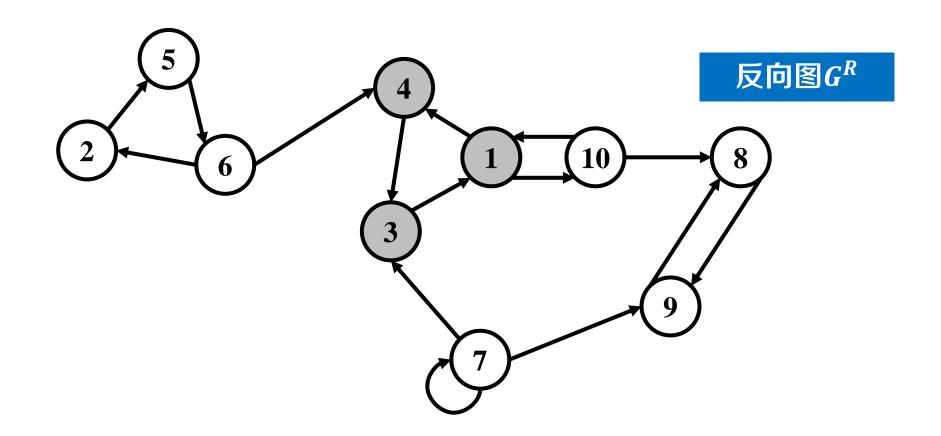


• 步骤1: 把边反向,得到反向图 G^R



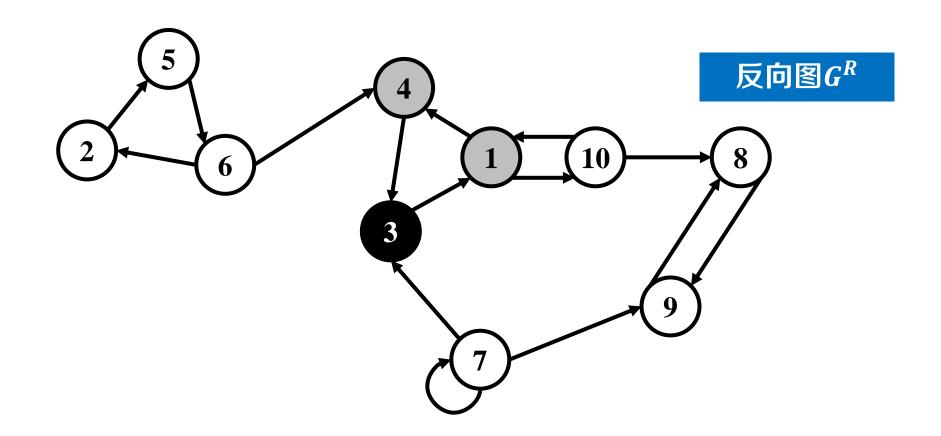


• 步骤1: 把边反向,得到反向图 G^R



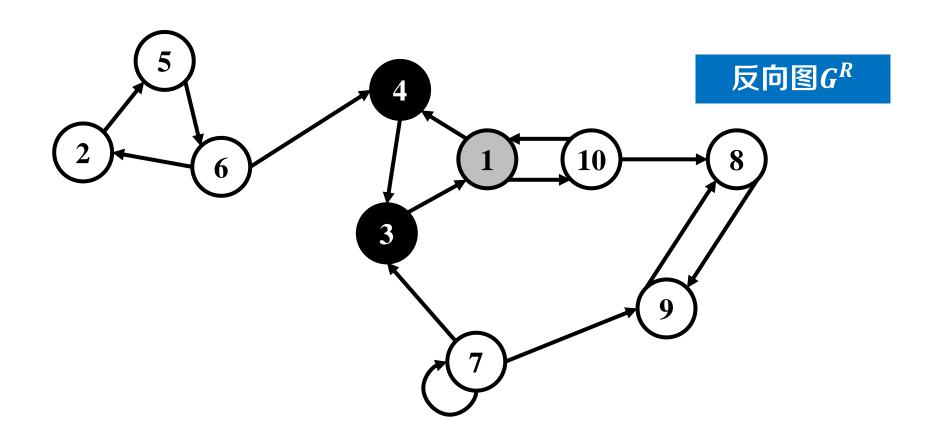


• 步骤1: 把边反向,得到反向图 G^R



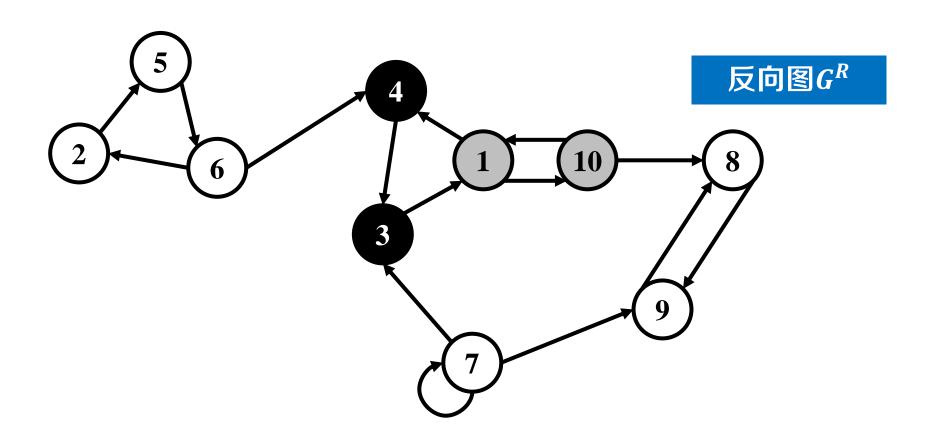


• 步骤1: 把边反向,得到反向图 G^R



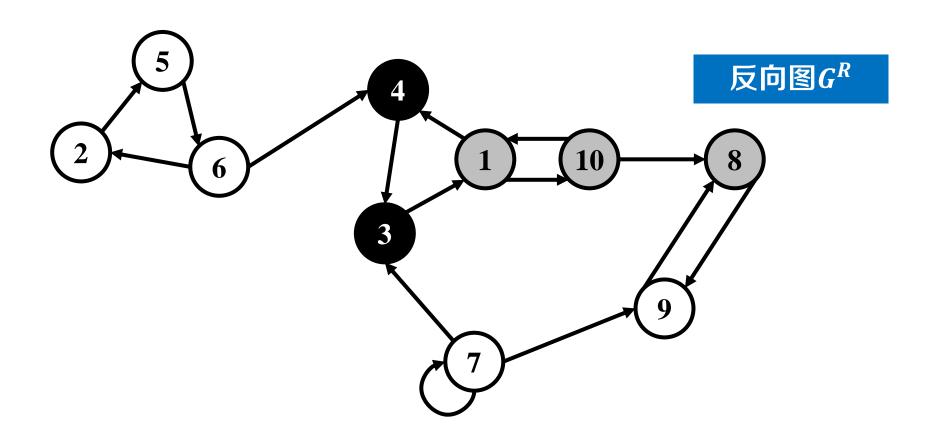


• 步骤1: 把边反向,得到反向图 G^R



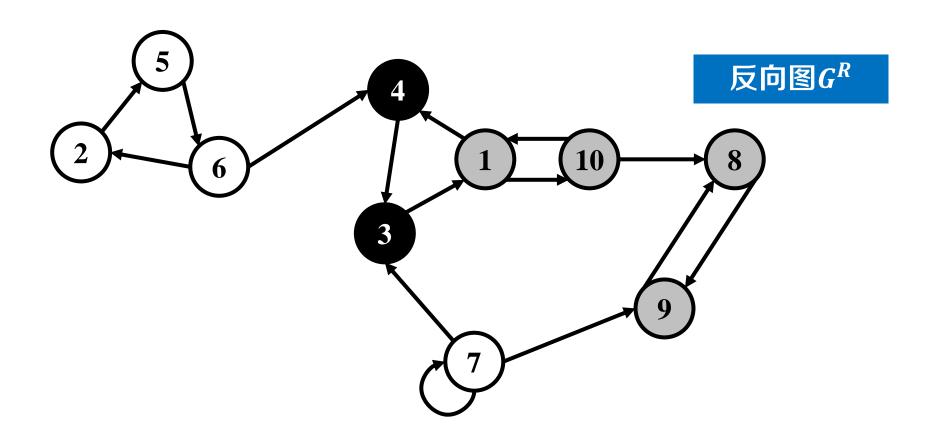


• 步骤1: 把边反向,得到反向图 G^R



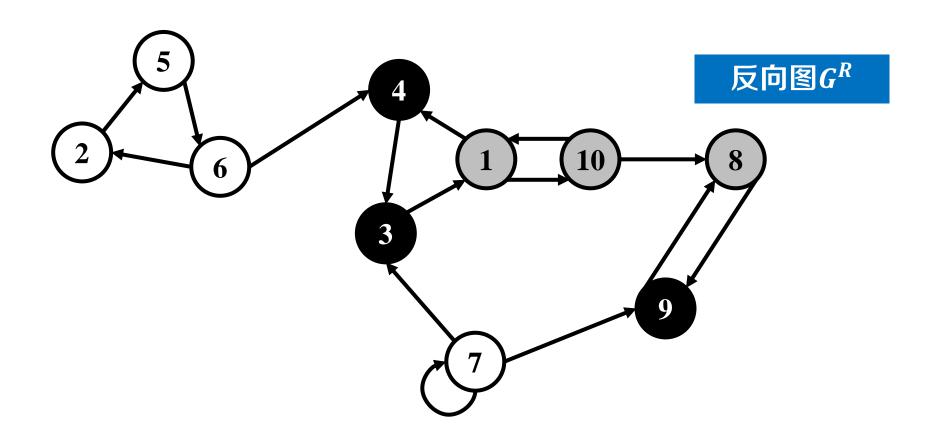


• 步骤1: 把边反向,得到反向图 G^R



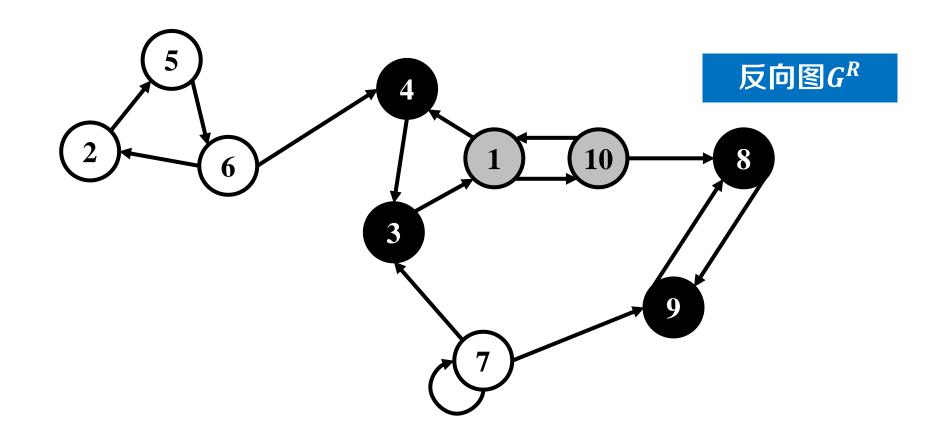


• 步骤1: 把边反向,得到反向图 G^R



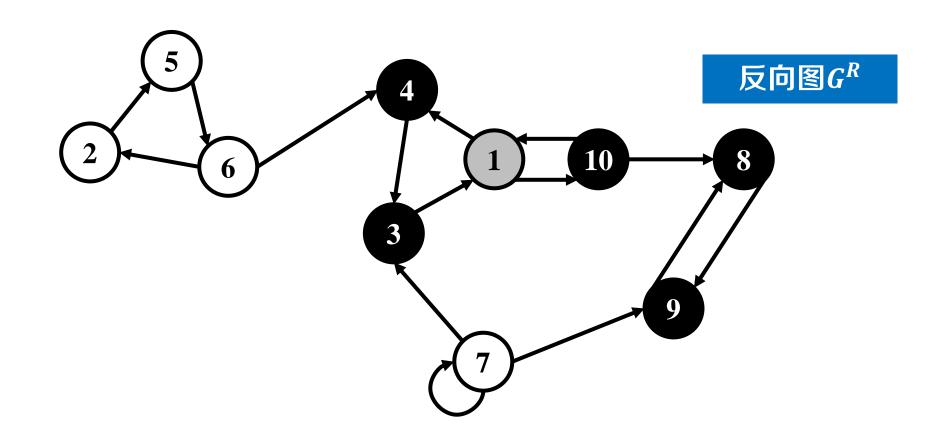


• 步骤1: 把边反向,得到反向图 G^R





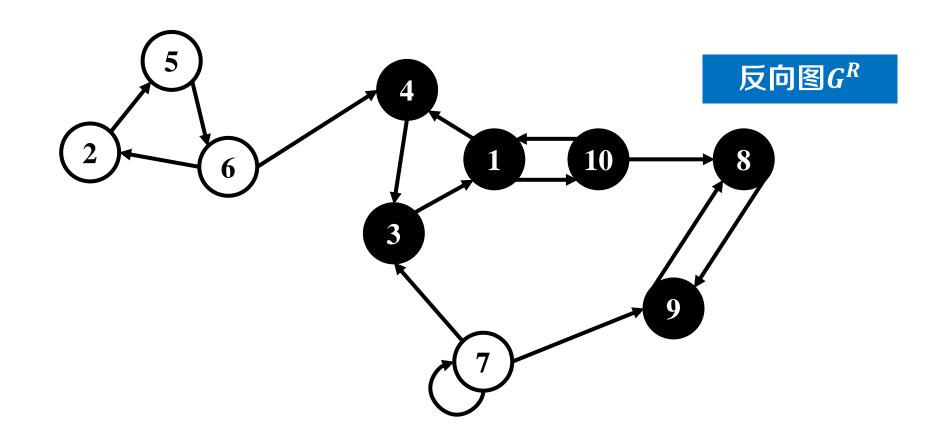
• 步骤1: 把边反向,得到反向图 G^R





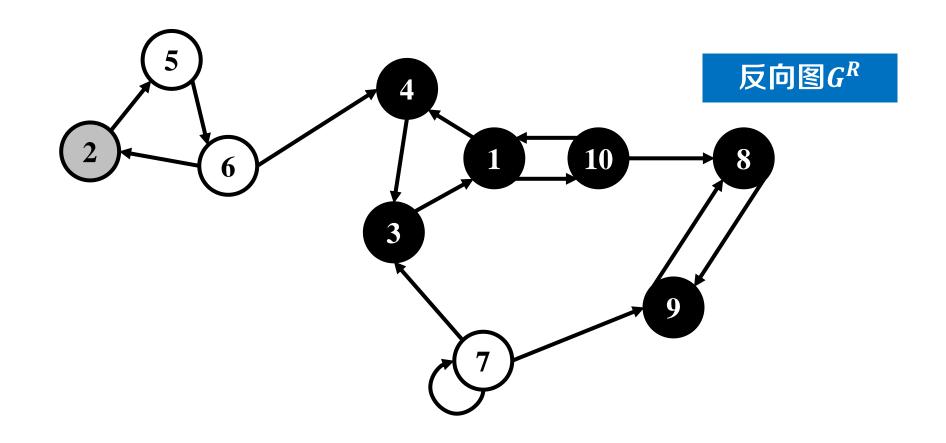
• 步骤1: 把边反向,得到反向图 G^R

10



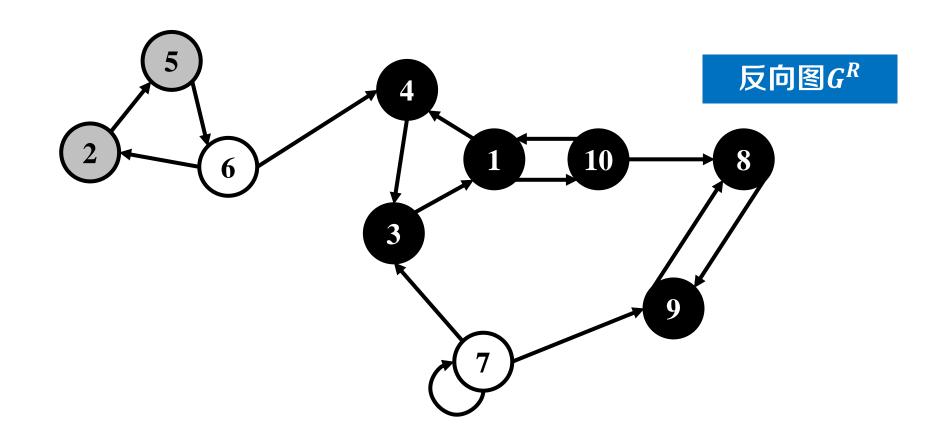


• 步骤1: 把边反向,得到反向图 G^R



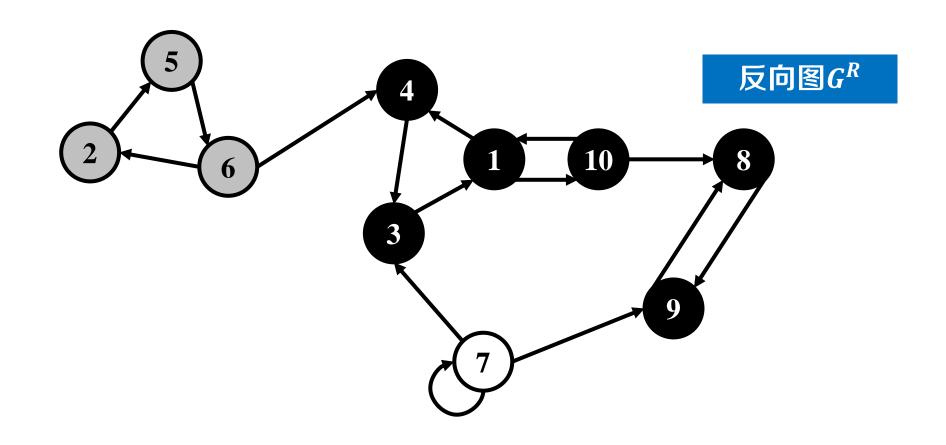


• 步骤1: 把边反向,得到反向图 G^R



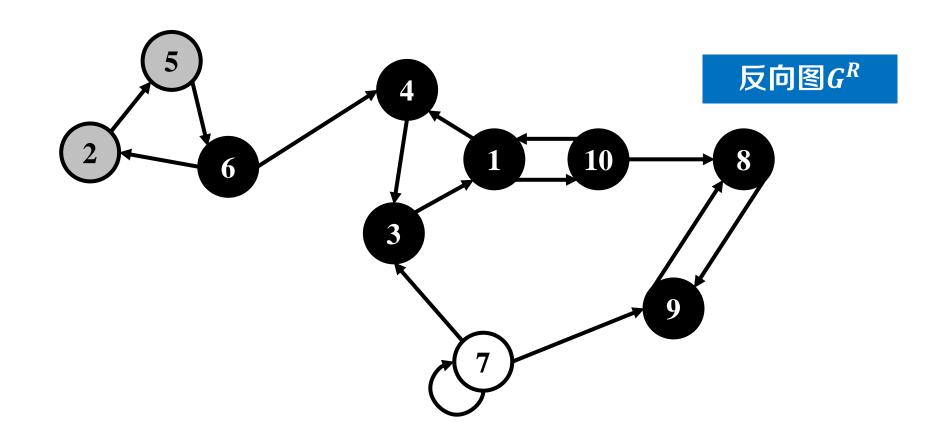


• 步骤1: 把边反向,得到反向图 G^R



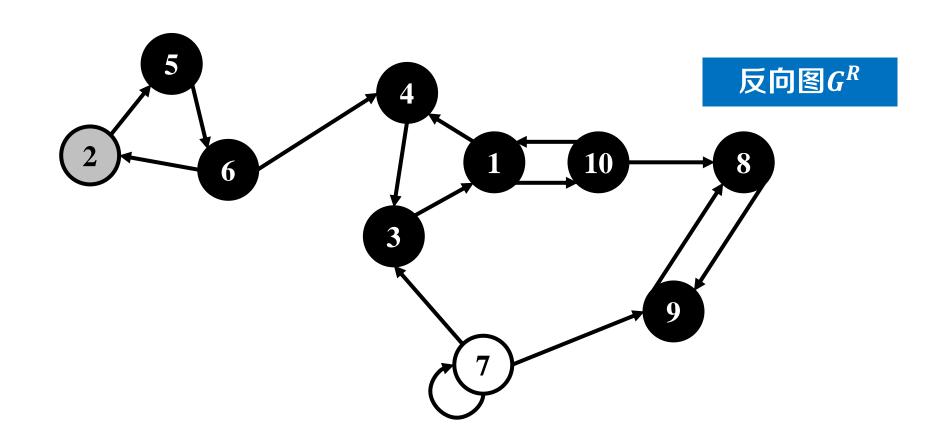


• 步骤1: 把边反向,得到反向图 G^R



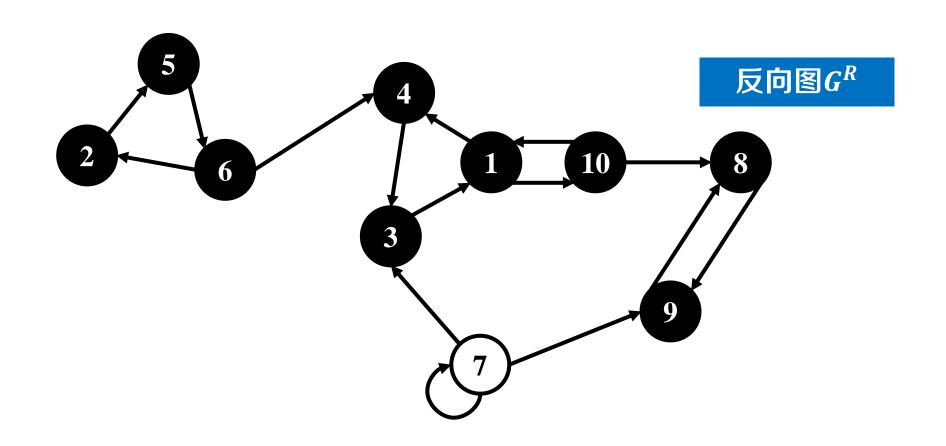


• 步骤1: 把边反向,得到反向图 G^R



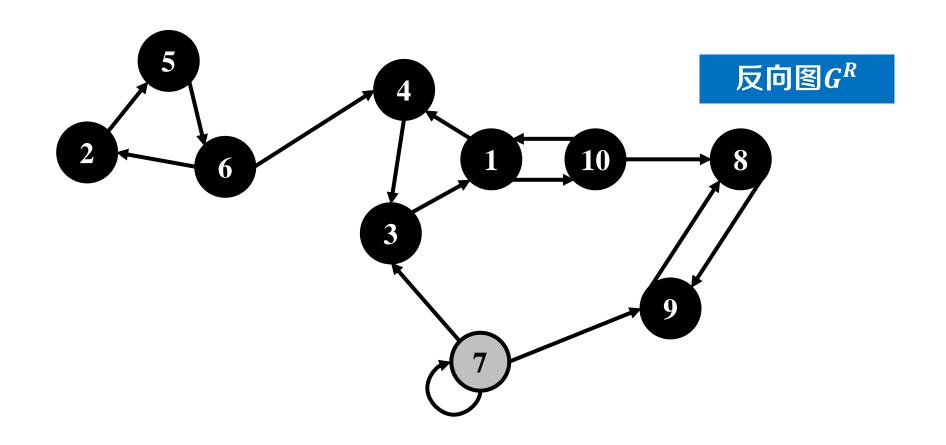


• 步骤1: 把边反向,得到反向图 G^R





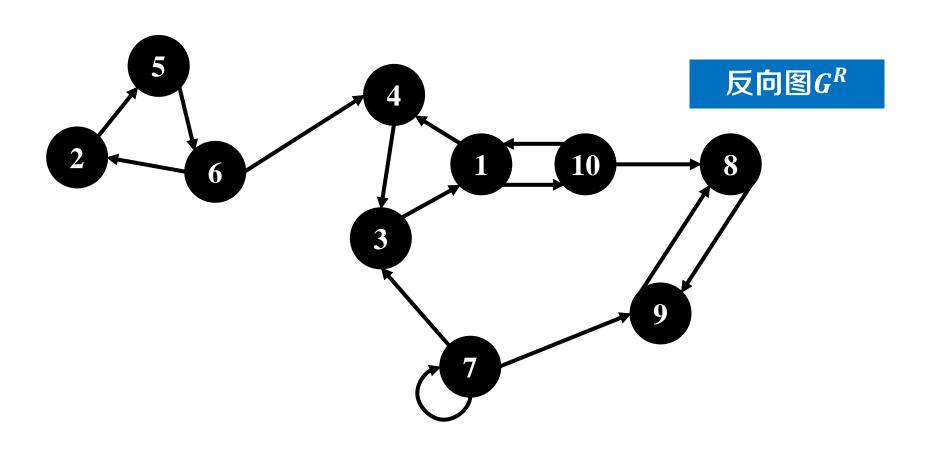
• 步骤1: 把边反向,得到反向图 G^R





• 步骤1: 把边反向,得到反向图 G^R

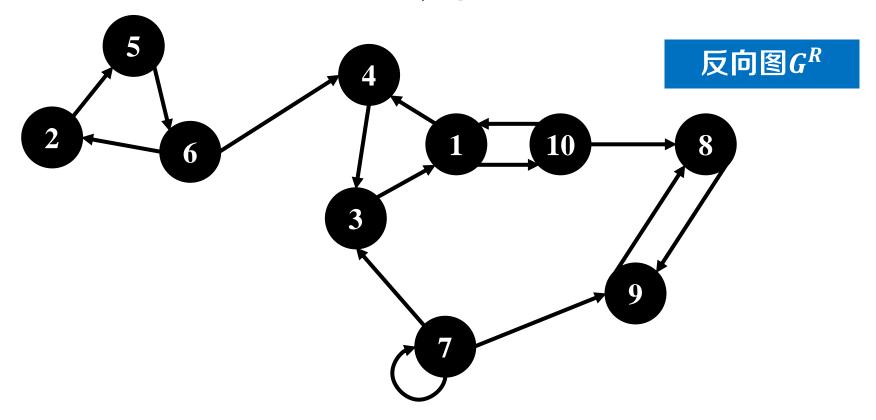
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

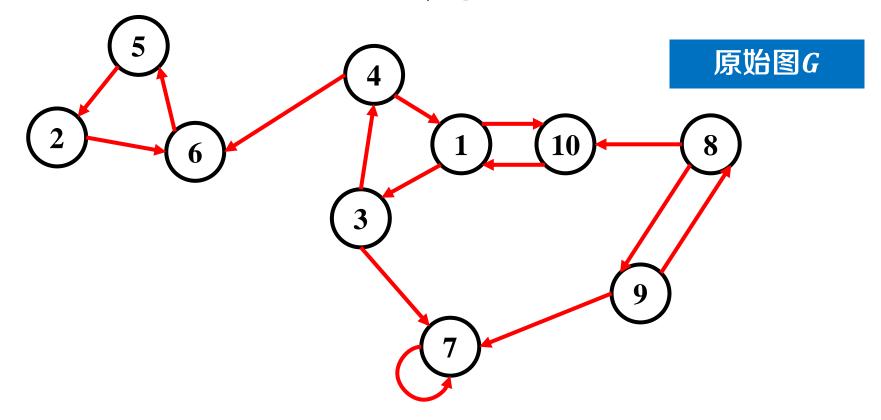
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

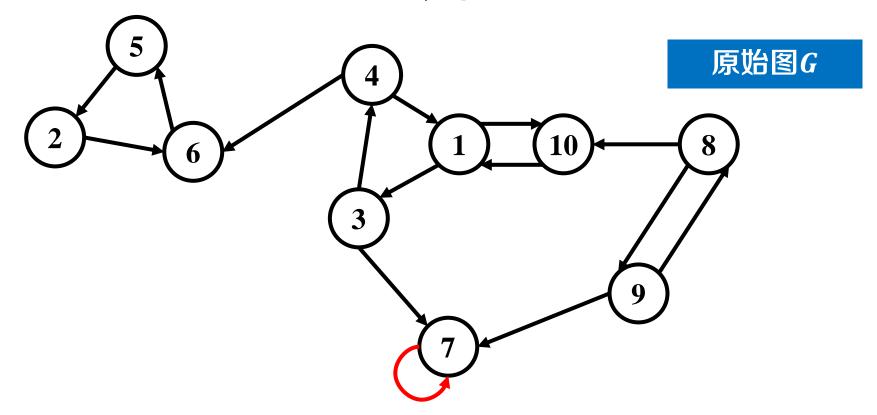
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

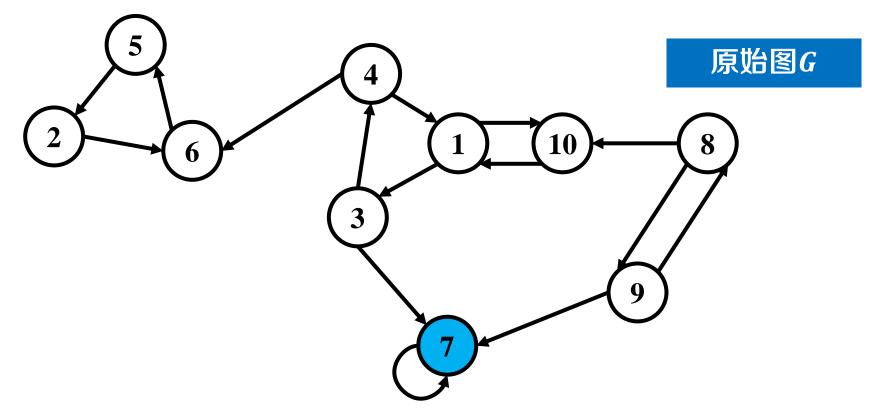
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

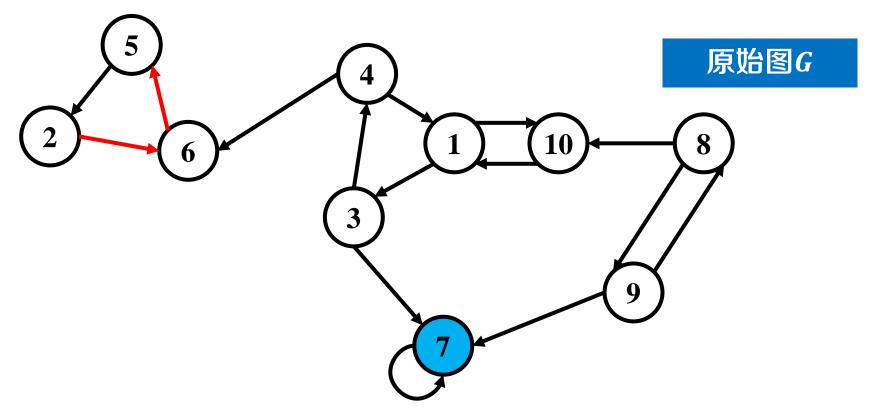
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

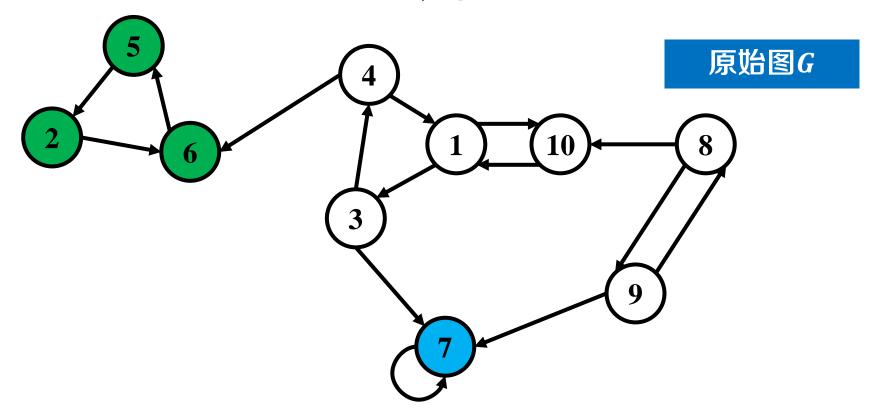
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

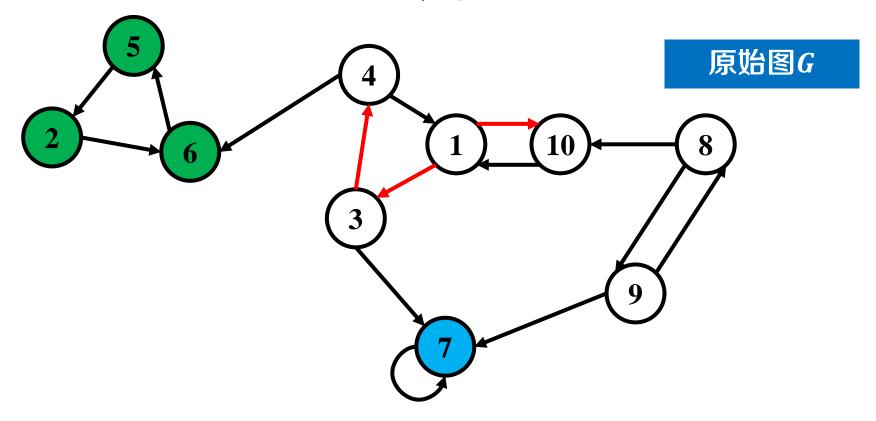
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

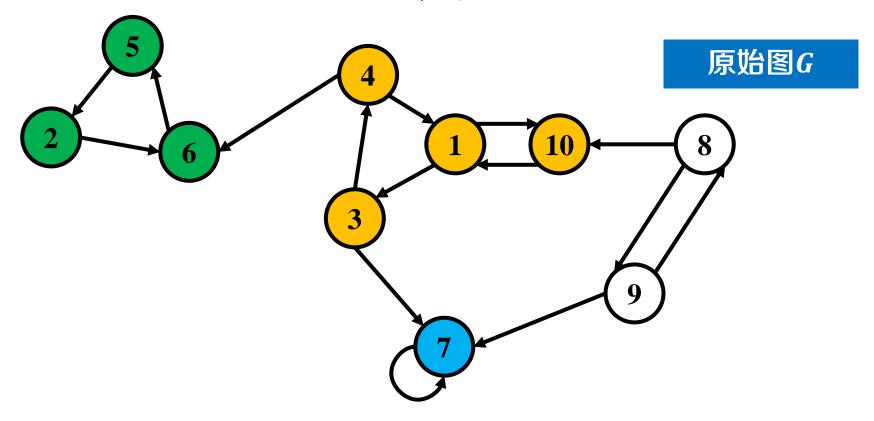
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

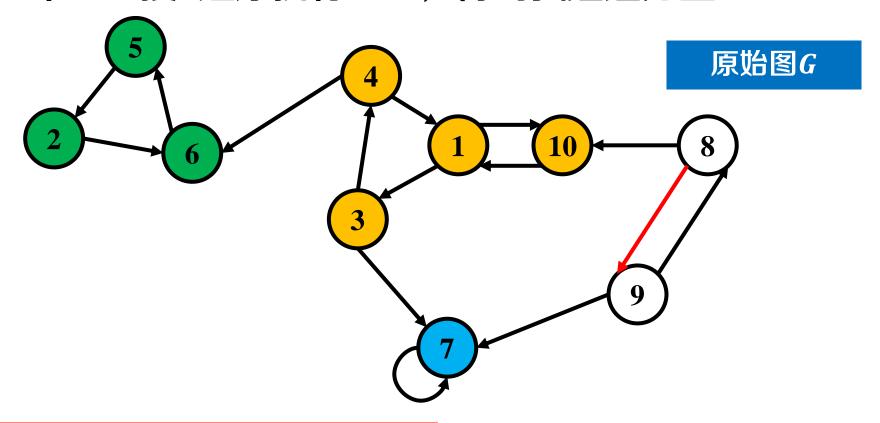
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

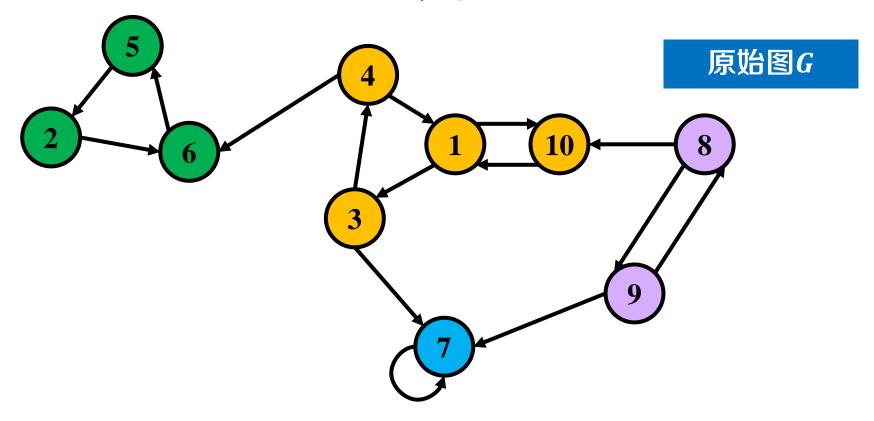
• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





• 步骤1: 把边反向,得到反向图 G^R

• 步骤2:在 G^R 上执行DFS,得到顶点完成时刻顺序L





问题背景与定义

算法框架与实例

伪代码与复杂度

算法正确性证明



```
输入: 图 G
   输出: 强连通分量
 \begin{array}{l} B \leftarrow \{\} \\ G^R \leftarrow G.reverse() \\ \overline{L} \leftarrow \mathrm{DFS}(G^R) \end{array} 
                                                                               构造反向图
   color[1..V] \leftarrow WHITE
   for i \leftarrow L.length() downto 1 do
        u \leftarrow L[i]
        if color[u] = WHITE then
              L_{scc} \leftarrow \text{DFS-Visit}(G, u)
            R \leftarrow R \cup set(L_{scc})
         end
   end
   return R
```



```
输入: 图 G
输出: 强连通分量
R \leftarrow \{\}
G^R \leftarrow G.reverse() 
L \leftarrow DFS(G^R)
                                                     在反向图上执行DFS
color[1..V] \leftarrow WHITE
for i \leftarrow L.length() downto 1 do
    u \leftarrow L[i]
    if color[u] = WHITE then
        L_{scc} \leftarrow \text{DFS-Visit}(G, u)
       R \leftarrow R \cup set(L_{scc})
    end
end
return R
```



```
输入: 图 G
输出: 强连通分量
R \leftarrow \{\}
G^R \leftarrow G.reverse()
L \leftarrow \mathrm{DFS}(G^R)
color[1..V] \leftarrow WHITE
for i \leftarrow L.length() downto 1 do
                                                   按L逆序在原图执行DFS
     u \leftarrow L[i]
    if color[u] = WHITE then
         L_{scc} \leftarrow \text{DFS-Visit}(G, u)
        R \leftarrow R \cup set(L_{scc})
     end
end
return R
```



```
输入: 图 G
  输出: 强连通分量
  R \leftarrow \{\}
G^R \leftarrow G.reverse() 
L \leftarrow DFS(G^R)
                                                   如何在搜索过程中得到L?
  color[1..V] \leftarrow WHITE
  for i \leftarrow L.length() downto 1 do
      u \leftarrow L[i]
      if color[u] = WHITE then
          L_{scc} \leftarrow \text{DFS-Visit}(G, u)
         R \leftarrow R \cup set(L_{scc})
      end
  \mathbf{end}
  return R
```

伪代码



• **DFS**(*G*)

```
输入: 图 G
新建数组 color[1..V], L[1..V]
for v \in V do
 | color[v] \leftarrow WHITE
end
for v \in V do
   if color[v] = WHITE then
      L' \leftarrow \text{DFS-Visit}(G, v)
       向L结尾追加L'
   end
end
return L
```

• DFS-Visit(G, v)

```
输入: 图 G, 顶点 v
 输出: 按完成时刻从早到晚排列的顶点 L
color[v] \leftarrow GRAY
for w \in G.Adj[v] do
  \mathbf{if} \ \underline{color}[w] = WHITE \ \mathbf{then}
     L \leftarrow \text{DFS-Visit}(G, w)
                                 顶点按
    end
                                 完成时
end
                                 刻排列
color[v] \leftarrow BLACK
向L结尾追加顶点v
return L
```

复杂度分析



```
输入: 图 G
输出: 强连通分量
R \leftarrow \{\}
G^R \leftarrow G.reverse() -----O(|V|+|E|)
L \leftarrow \mathrm{DFS}(G^R)
                       ----O(|V|+|E|) 第一次深度优先搜索
color[1..V] \leftarrow WHITE
for i \leftarrow L.length() downto 1 do
   u \leftarrow L[i]
   if color[u] = WHITE then
                                   O(|V| + |E|) 第二次深度优先搜索
      L_{scc} \leftarrow \text{DFS-Visit}(G, u)
     R \leftarrow R \cup set(L_{scc})
   end
end
                                             时间复杂度: O(|V| + |E|)
return R
```



问题背景与定义

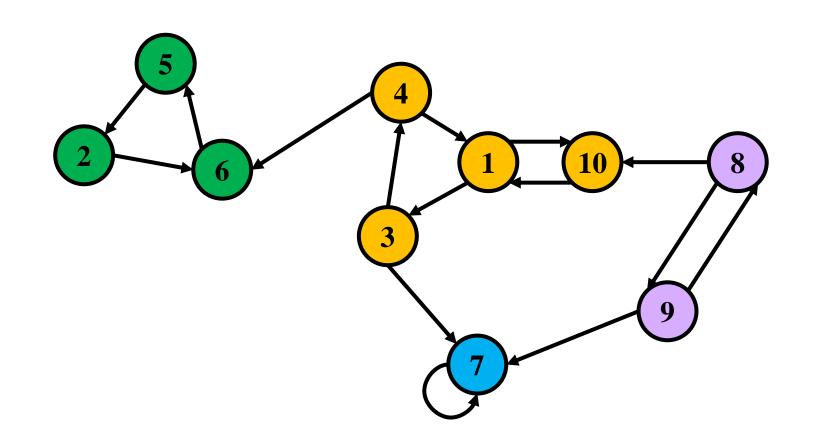
算法框架与实例

伪代码与复杂度

算法正确性证明

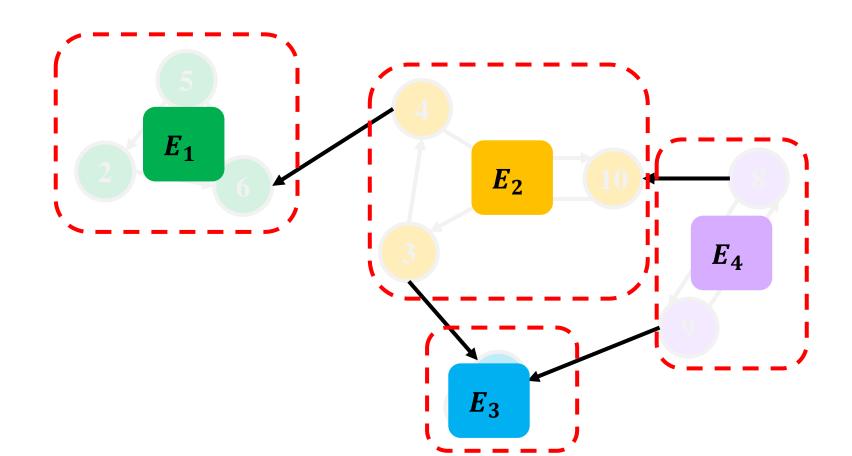


• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图



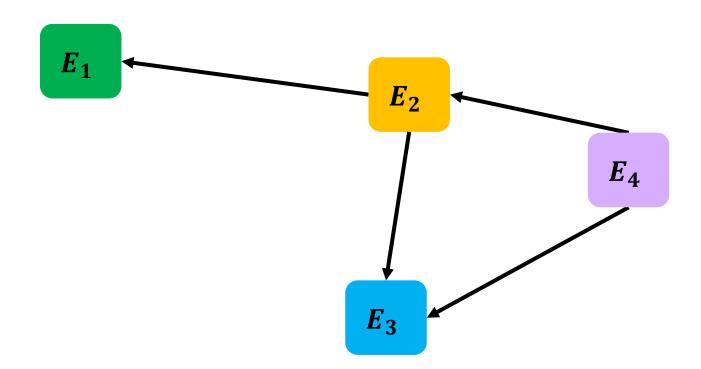


• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图





• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图



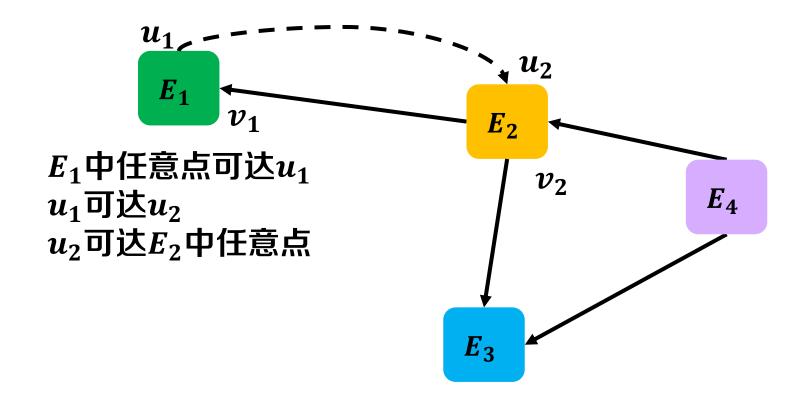


• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图

● 性质: *G^{SCC}*一定是有向无环图

反证:若存在环,两强连通分量中顶点相互可达,与最大性矛盾。

最大性: 加入新顶点,不保证相互可达

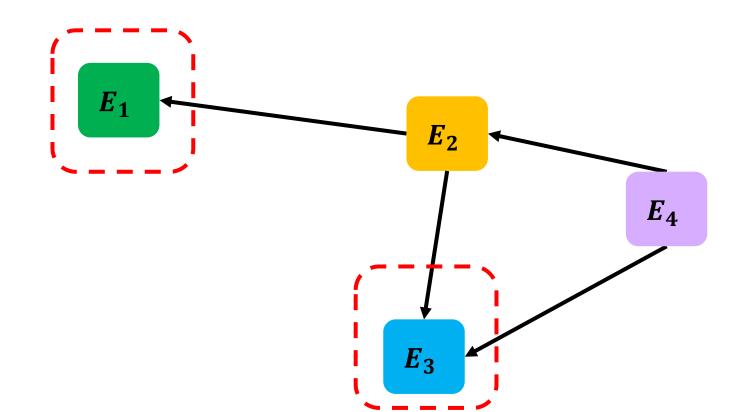




• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图

• 性质: *GSCC*一定是有向无环图

• SCC_{Sink}: G^{SCC}中出度为0的点





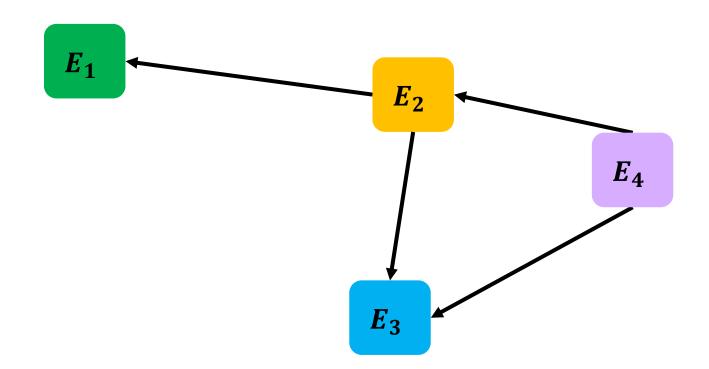
• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图

• 性质: *GSCC*一定是有向无环图

• SCC_{Sink}: G^{SCC}中出度为0的点

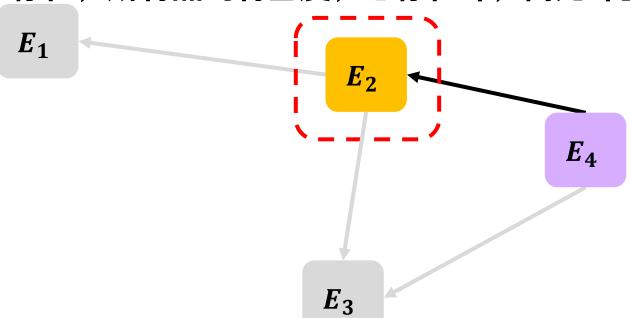
● 性质1: G^{SCC}中存在至少一个SCC_{Sink}

• 反证: 若不存在,所有点均有出度,必存在环,矛盾





- 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图
 - 性质: *GSCC*一定是有向无环图
- SCC_{Sink}: G^{SCC}中出度为0的点
 - 性质1: G^{SCC} 中存在至少一个 SCC_{Sink}
 - 性质2: 删除 SCC_{Sink} , 会产生新的 SCC_{Sink}
 - 反证:若不存在,所有点均有出度,必存在环;而无环图子图必无环,矛盾





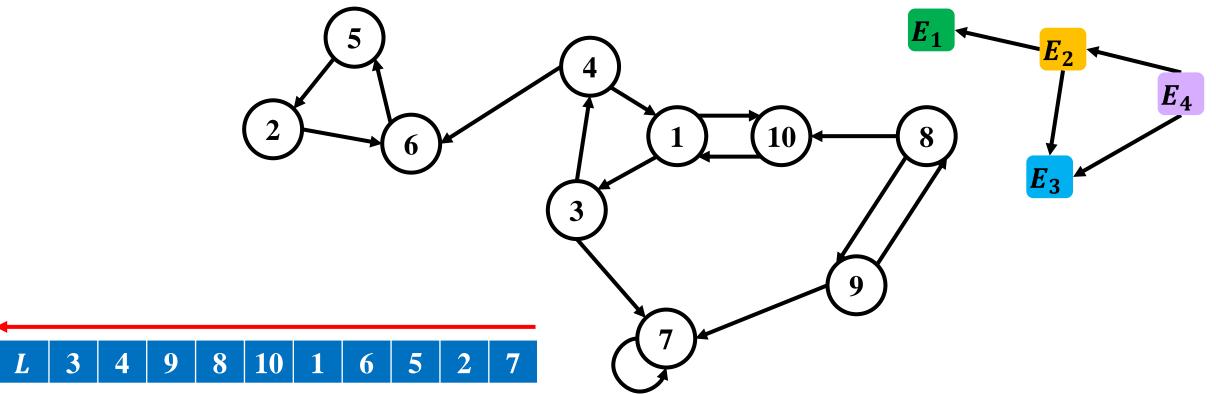
• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图

• 性质: *GSCC*一定是有向无环图

• SCC_{Sink}: G^{SCC}中出度为0的点

● 性质1: G^{SCC}中存在至少一个SCC_{Sink}

• 性质2: 删除 SCC_{Sink} , 会产生新的 SCC_{Sink}





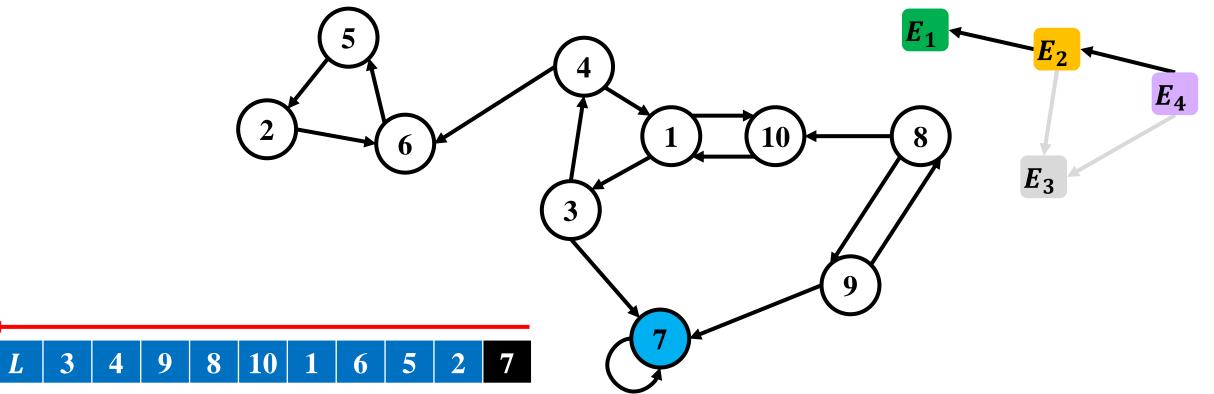
• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图

• 性质: *G^{SCC}*一定是有向无环图

• SCC_{Sink}: G^{SCC}中出度为0的点

● 性质1: G^{SCC}中存在至少一个SCC_{Sink}

• 性质2: 删除 SCC_{Sink} , 会产生新的 SCC_{Sink}





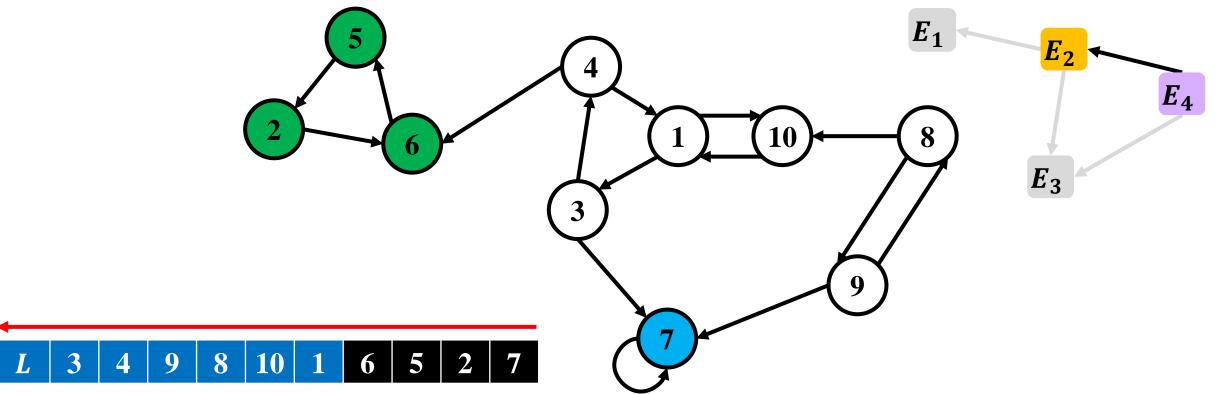
• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图

• 性质: *G^{SCC}*一定是有向无环图

• SCC_{Sink}: G^{SCC}中出度为0的点

● 性质1: G^{SCC}中存在至少一个SCC_{Sink}

• 性质2: 删除 SCC_{Sink} , 会产生新的 SCC_{Sink}





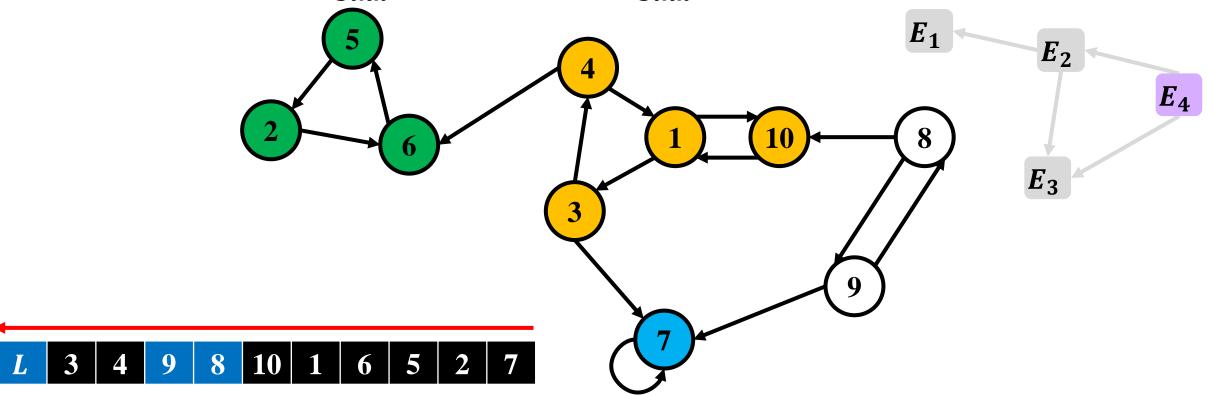
• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图

• 性质: *GSCC*一定是有向无环图

• SCC_{Sink}: G^{SCC}中出度为0的点

● 性质1: G^{SCC}中存在至少一个SCC_{Sink}

• 性质2: 删除 SCC_{Sink} , 会产生新的 SCC_{Sink}





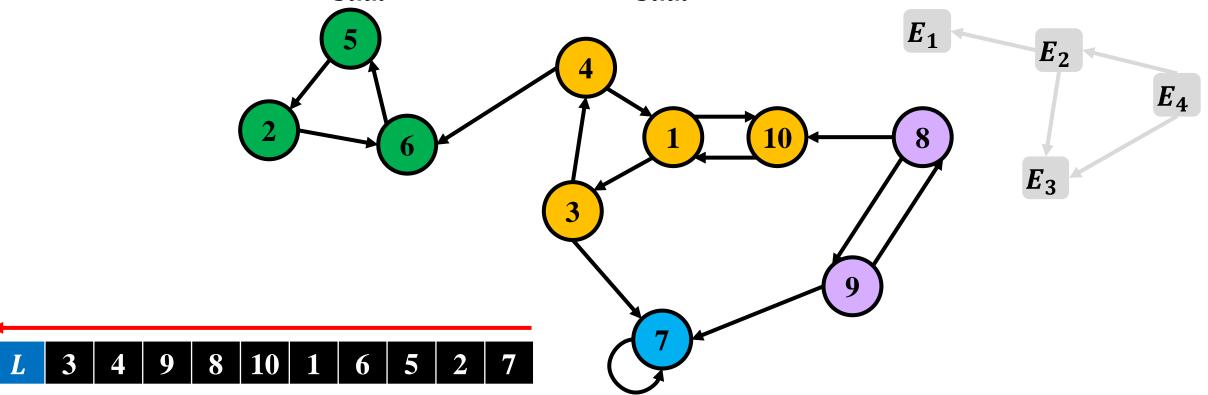
• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图

• 性质: *G^{SCC}*一定是有向无环图

• SCC_{Sink}: G^{SCC}中出度为0的点

● 性质1: G^{SCC}中存在至少一个SCC_{Sink}

• 性质2: 删除 SCC_{Sink} , 会产生新的 SCC_{Sink}





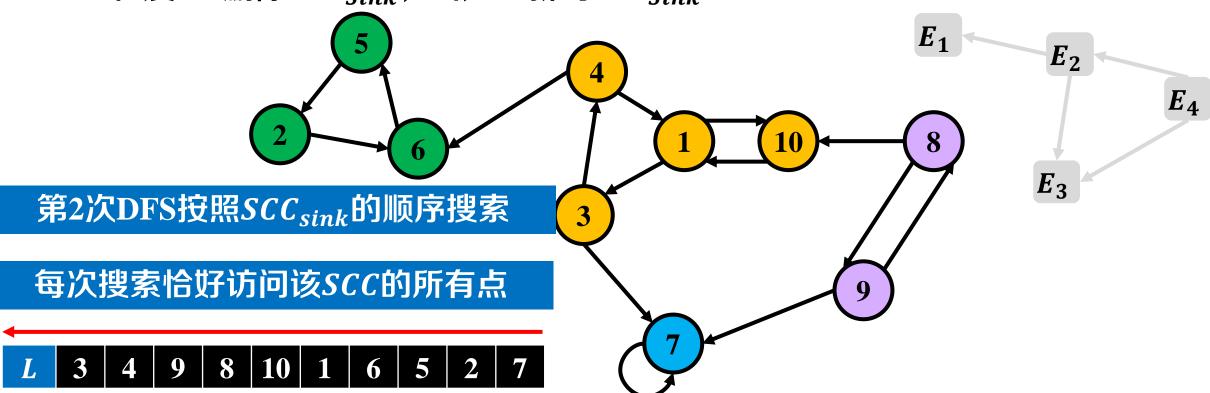
• 强连通分量图 G^{SCC} : 把强连通分量看作一个点,得到有向图

• 性质: *G^{SCC}*一定是有向无环图

• SCC_{Sink}: G^{SCC}中出度为0的点

• 性质1: G^{SCC} 中存在至少一个 SCC_{Sink}

• 性质2: 删除 SCC_{Sink} , 会产生新的 SCC_{Sink}





通过按L逆序执行DFS实现



第2次DFS按照SCCsink的顺序搜索



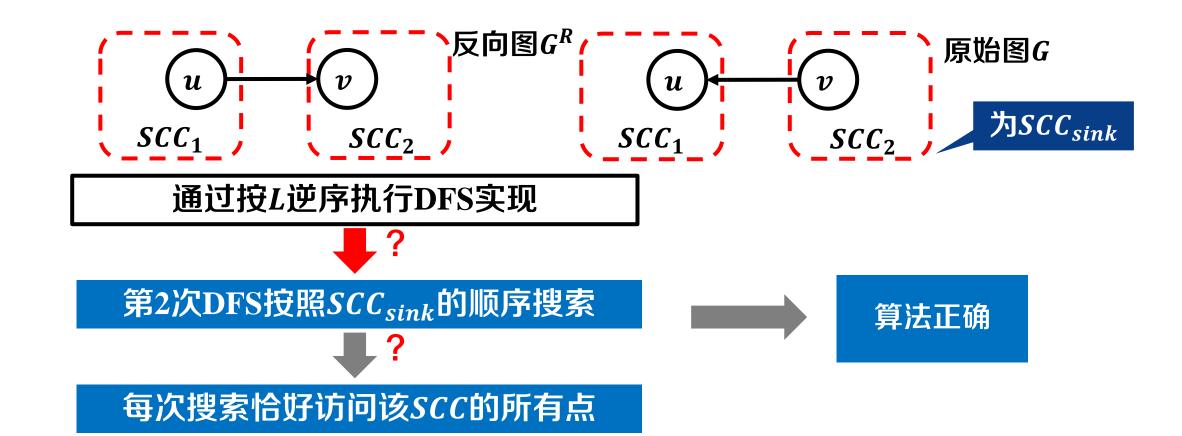
每次搜索恰好访问该SCC的所有点



算法正确

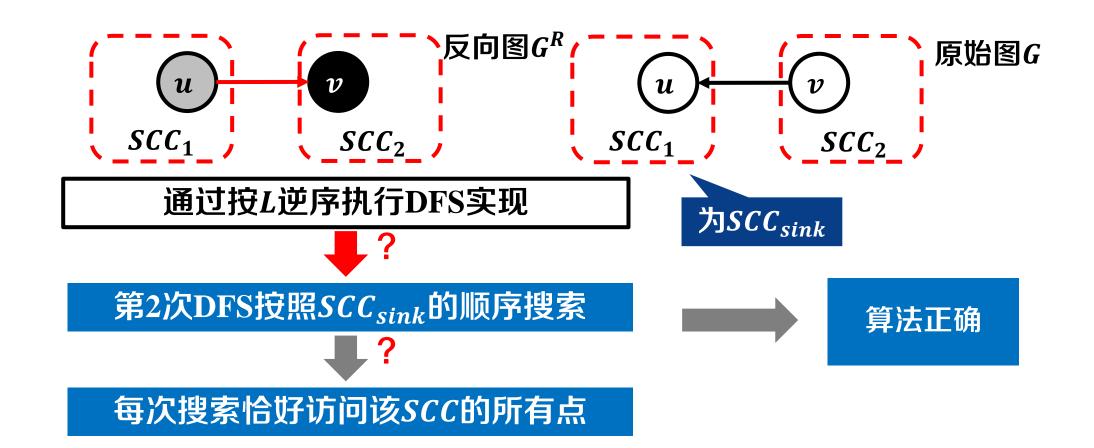


• 给定反向图 G^R ,存在边(u,v), $u \in SCC_1$, $v \in SCC_2$



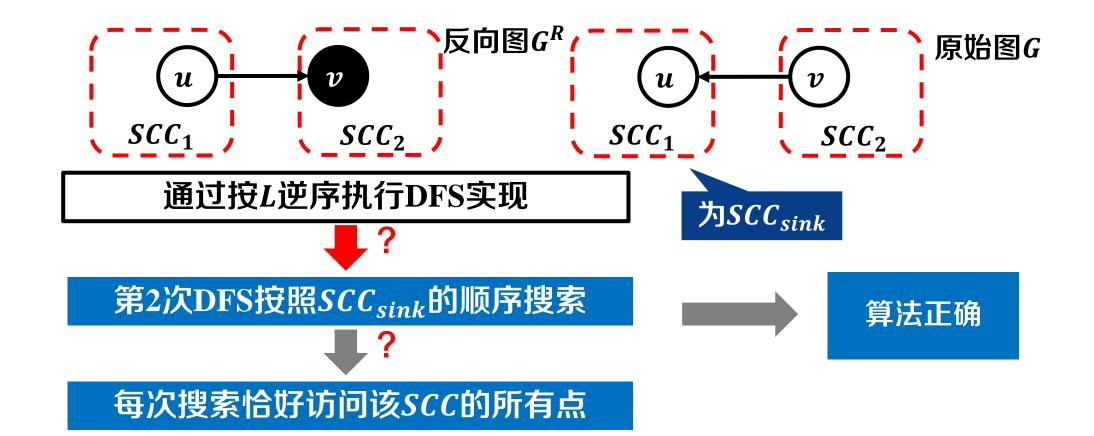


- 给定反向图 G^R ,存在边(u,v), $u \in SCC_1$, $v \in SCC_2$
 - 若先搜索u,会从u搜索v,则f(v) < f(u)





- 给定反向图 G^R ,存在边(u,v), $u \in SCC_1$, $v \in SCC_2$
 - 若先搜索u,会从u搜索v,则f(v) < f(u)
 - 若先搜索v, v搜索完成才开始搜索u, 则f(v) < f(u)



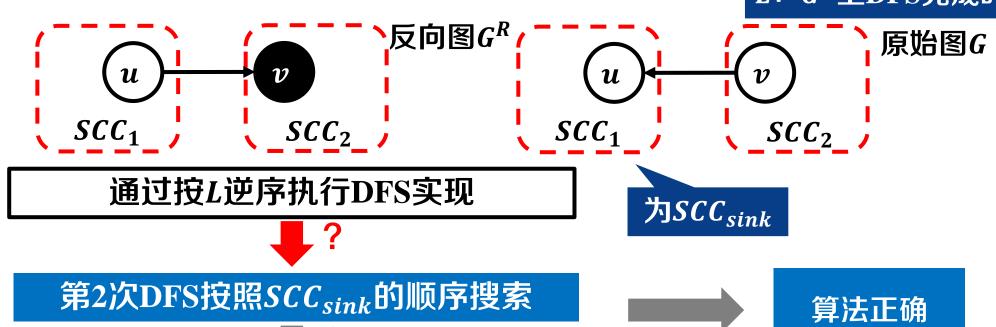


- 给定反向图 G^R ,存在边(u,v), $u \in SCC_1$, $v \in SCC_2$
 - f越来越小

- 若先搜索u,会从u搜索v,则f(v) < f(u)
- 若先搜索v, v搜索完成才开始搜索u, 则f(v) < f(u)



L: GR上DFS完成时刻顺序





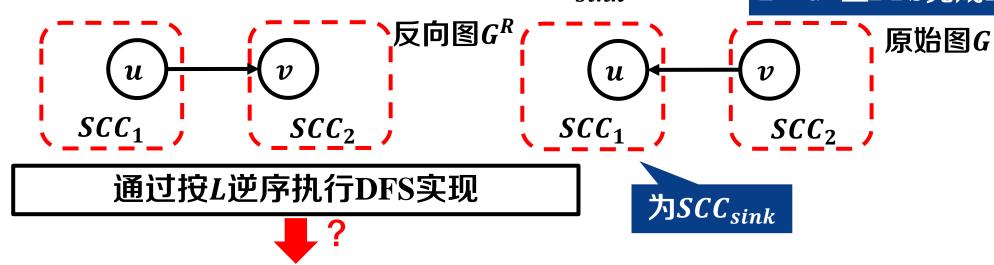
每次搜索恰好访问该SCC的所有点



- 给定反向图 G^R ,存在边(u,v), $u \in SCC_1$, $v \in SCC_2$
 - f越来越小

- 若先搜索u,会从u搜索v,则f(v) < f(u)
- 若先搜索v, v搜索完成才开始搜索u, 则f(v) < f(u)
- 所以按L的逆序,总是先搜索u,符合 SCC_{sink} 的顺序

 $L: G^R \perp DFS$ 完成时刻顺序



第2次DFS按照SCCsink的顺序搜索



算法正确

每次搜索恰好访问该SCC的所有点

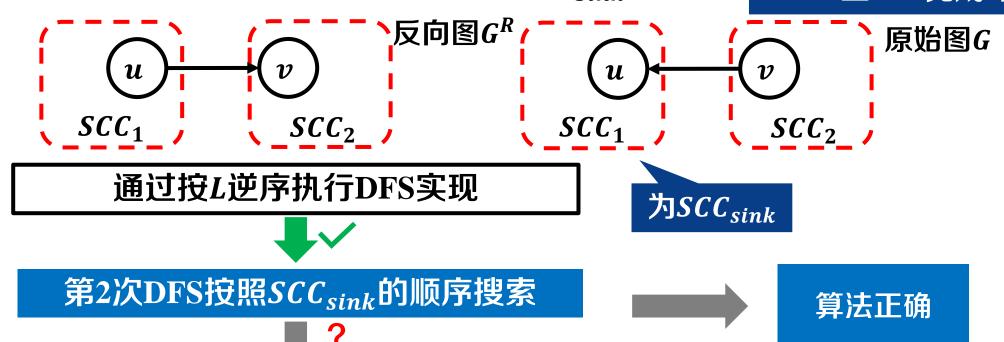


- 给定反向图 G^R ,存在边(u,v), $u \in SCC_1$, $v \in SCC_2$ _{fix}
 - f越来越小

- 若先搜索u,会从u搜索v,则f(v) < f(u)
- 若先搜索v, v搜索完成才开始搜索u, 则f(v) < f(u)
- 所以按L的逆序,总是先搜索u,符合 SCC_{sink} 的顺序

L ... v ... u ...

 $L: G^R \perp DFS$ 完成时刻顺序



每次搜索恰好访问该SCC的所有点

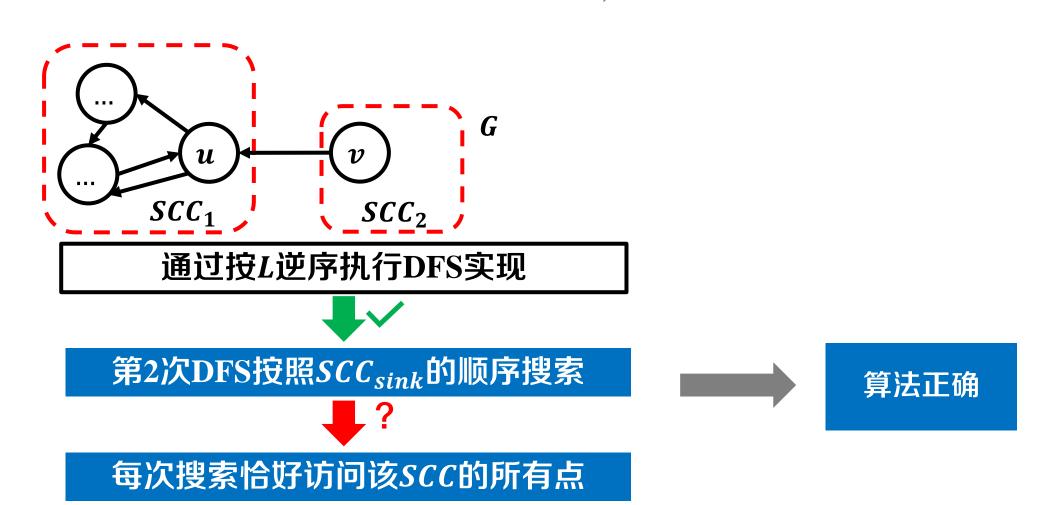


• 强连通分量内,顶点相互可达

DFS可以访问到该SCC所有点

● SCC_{sink}出度为0

DFS不会访问该SCC以外的点



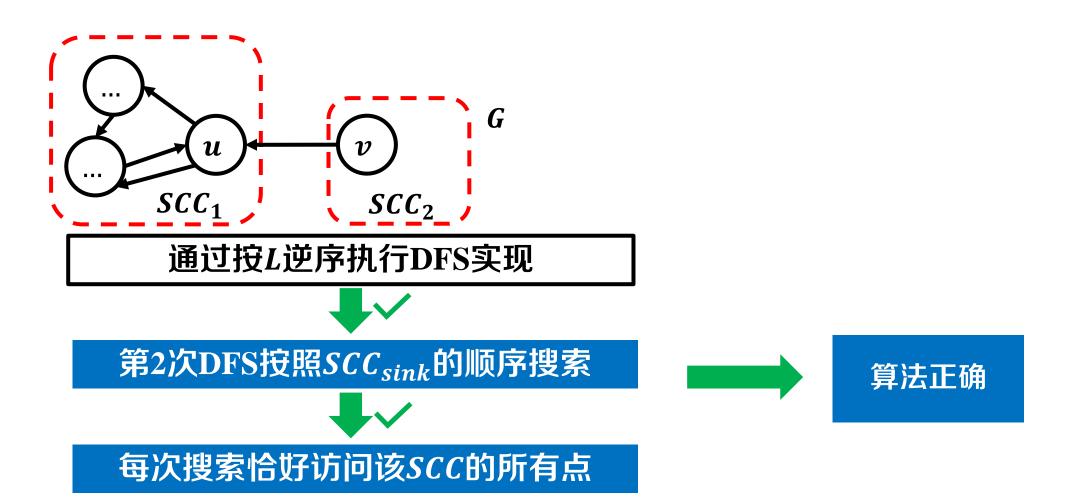


• 强连通分量内,顶点相互可达

DFS可以访问到该SCC所有点

● SCC_{sink}出度为0

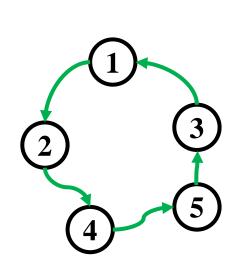
DFS不会访问该SCC以外的点

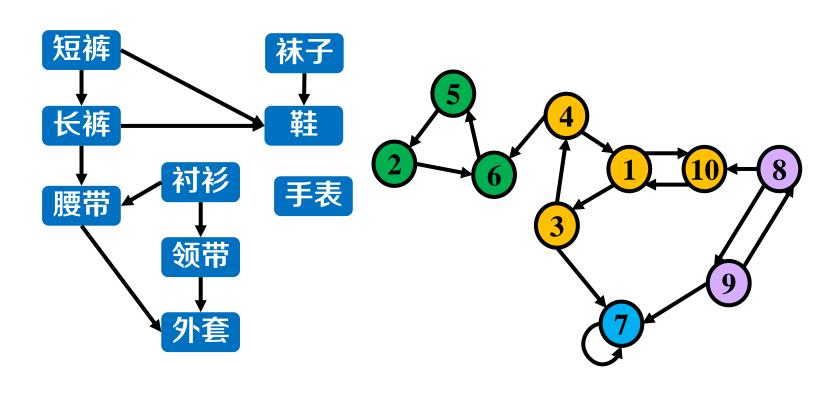


小结



• 深度优先搜索的应用





环的存在性判定

利用深度优先搜索边的性质

拓扑排序

强连通分量

利用深度优先搜索点的性质(括号化定理)





