

# Swing 召回通道

王树森

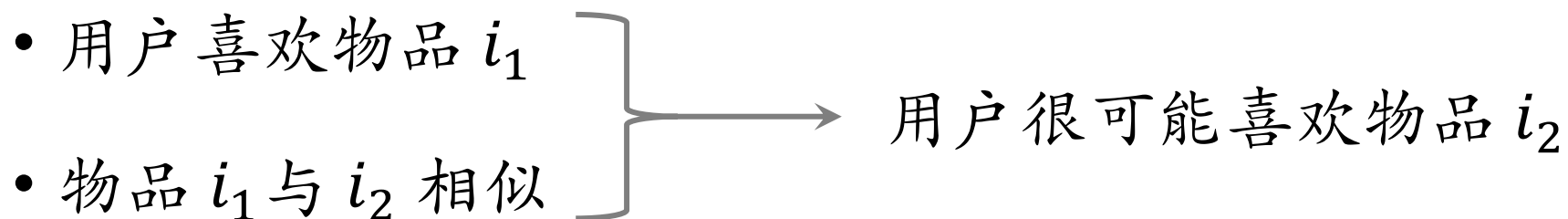
ShusenWang@xiaohongshu.com

<http://wangshusen.github.io/>



# ItemCF的原理

- 物品相似度：如果喜欢  $i_1$ 、 $i_2$  的用户有很大的重叠，那么  $i_1$  与  $i_2$  相似。



# ItemCF的物品相似度

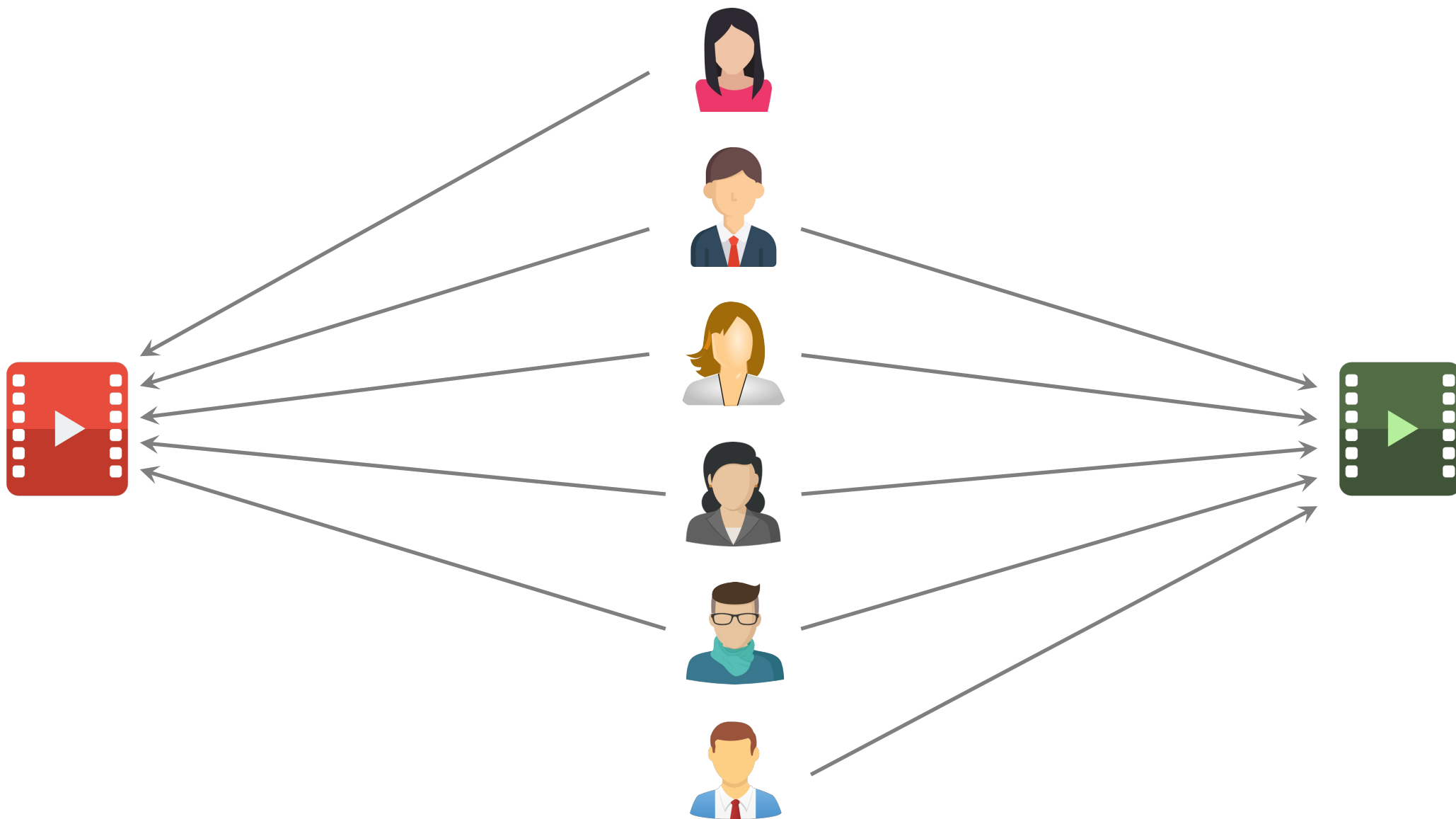
- 喜欢物品  $i_1$  的用户记作集合  $\mathcal{W}_1$ 。
- 喜欢物品  $i_2$  的用户记作集合  $\mathcal{W}_2$ 。
- 定义交集  $\mathcal{V} = \mathcal{W}_1 \cap \mathcal{W}_2$ 。

# ItemCF的物品相似度

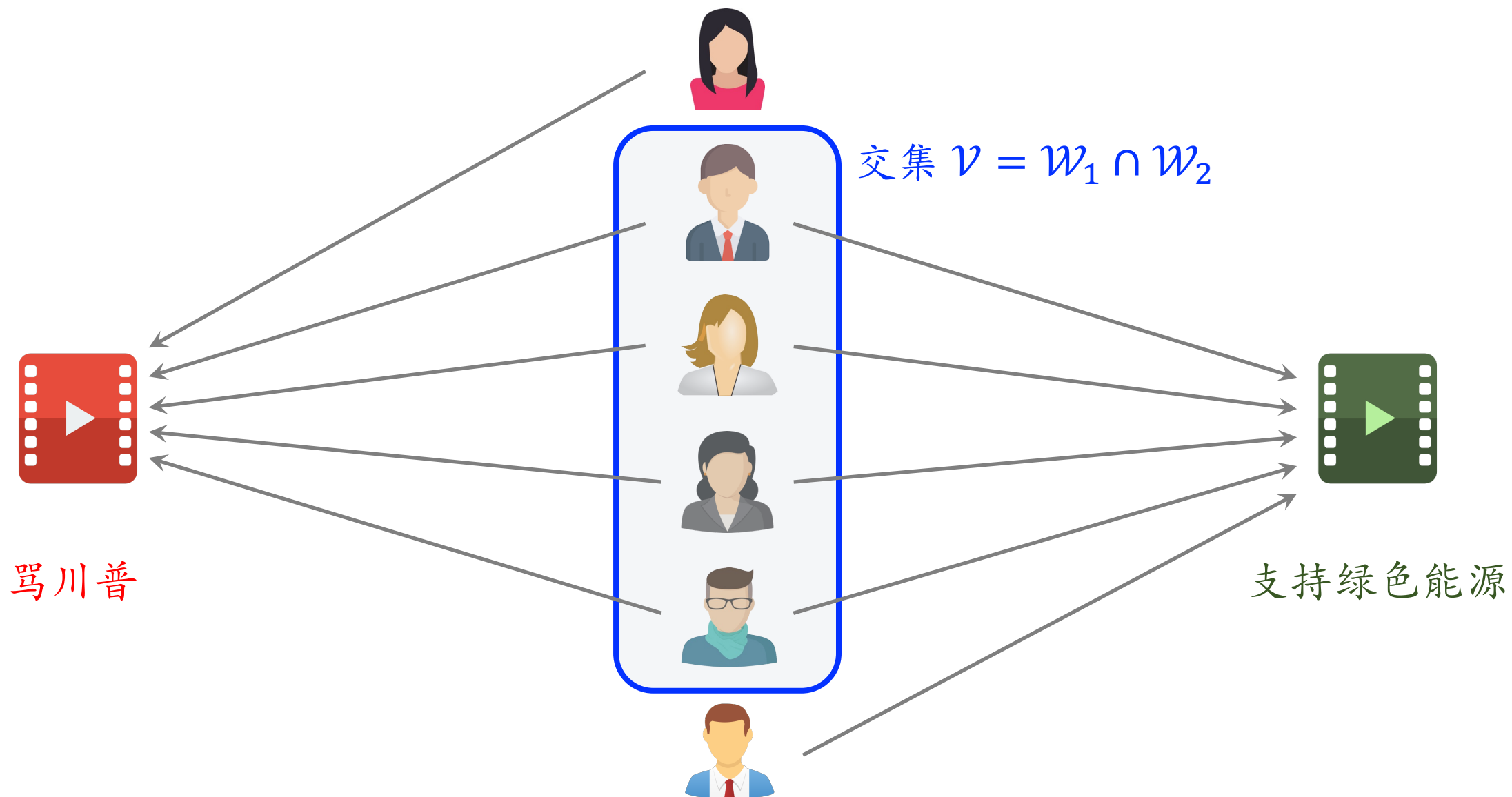
- 喜欢物品  $i_1$  的用户记作集合  $\mathcal{W}_1$ 。
- 喜欢物品  $i_2$  的用户记作集合  $\mathcal{W}_2$ 。
- 定义交集  $\mathcal{V} = \mathcal{W}_1 \cap \mathcal{W}_2$ 。
- 两个物品的相似度：

$$sim(i_1, i_2) = \frac{|\mathcal{V}|}{\sqrt{|\mathcal{W}_1| \cdot |\mathcal{W}_2|}}$$

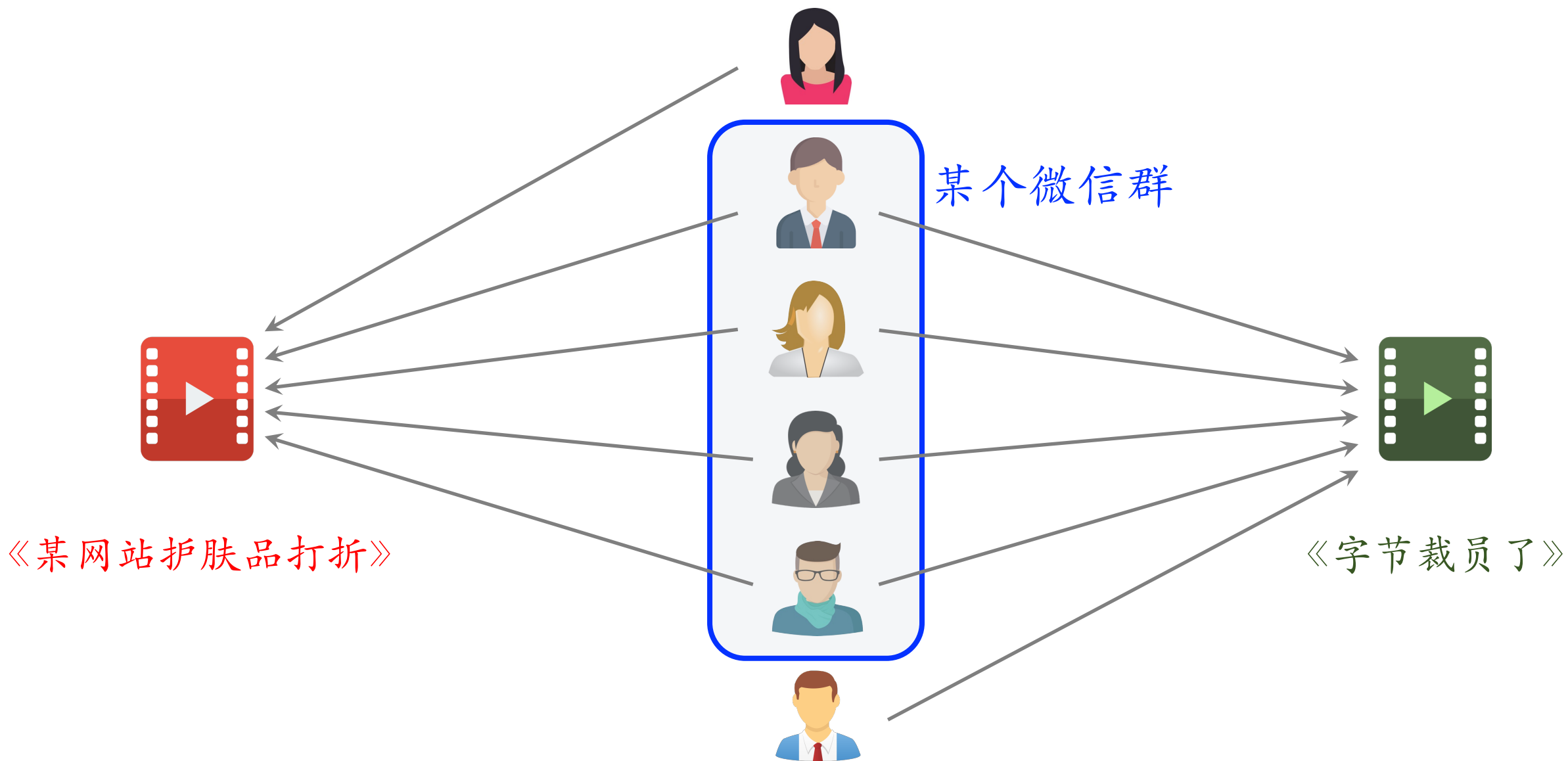
# ItemCF的物品相似度



# ItemCF的物品相似度



假如重合的用户是一个小圈子.....



# Swing模型

- 用户  $u_1$  喜欢的物品记作集合  $J_1$ 。
- 用户  $u_2$  喜欢的物品记作集合  $J_2$ 。
- 定义两个用户的重合度：

$$\text{overlap}(u_1, u_2) = |J_1 \cap J_2|。$$

- 用户  $u_1$  和  $u_2$  的重合度高，则他们可能来自一个小圈子，要降低他们的权重。



# Swing模型

- 喜欢物品  $i_1$  的用户记作集合  $\mathcal{W}_1$ 。
- 喜欢物品  $i_2$  的用户记作集合  $\mathcal{W}_2$ 。
- 定义交集  $\mathcal{V} = \mathcal{W}_1 \cap \mathcal{W}_2$ 。
- 两个物品的相似度：

$$\underline{sim(i_1, i_2)} = \sum_{u_1 \in \mathcal{V}} \sum_{u_2 \in \mathcal{V}} \frac{1}{\alpha + \text{overlap}(u_1, u_2)}$$

# 总结

- Swing 与 ItemCF 唯一的区别在于物品相似度。
- ItemCF：两个物品重合的用户比例高，则判定两个物品相似。
- Swing：额外考虑重合的用户是否来自一个小圈子。
  - 同时喜欢两个物品的用户记作集合  $\mathcal{V}$ 。
  - 对于  $\mathcal{V}$  中的用户  $u_1$  和  $u_2$ ，重合度记作  $\text{overlap}(u_1, u_2)$ 。
  - 两个用户重合度大，则可能来自一个小圈子，权重降低。

# 长期招聘优秀的算法工程师

- 部门：小红书社区技术部。
- 方向：搜索、推荐。
- 职位：校招、社招、实习。
- 地点：上海、北京。
- 联系方式：`ShusenWang@xiaohongshu.com`

**Thank You!**

<http://wangshusen.github.io/>