1.1 在课件中，我们给出了如下评分矩阵：

users

	1	2	3	4	**5**	6	7	8	9	10	11	12
1	1		3		?	5			5		4	
2			5	4			4			2	1	3
3	2	4		1	2		3		4	3	5	
4		2	4		5		4			2		
5			4	3	4	2					2	5
6	1		3		3		2			4		

movies

采用基于用户的评分预测方法（同样采用 2-最近邻），预测用户 5 对于电影 1 的评分，并与课件中给出的基于物品的评分结果进行比较。

各用户的平均打分：

$\overline{u_1} = (1+2+1)/3 = \frac{4}{3}$

$\overline{u_2} = (4+2)/2 = 3$

$\overline{u_3} = (3+5+4+4+3)/5 = \frac{19}{5} = 3.8$

$\overline{u_4} = (4+1+3)/3 = \frac{8}{3}$

$\overline{u_5} = (2+5+4+3)/4 = \frac{7}{2} = 3.5$

$\overline{u_6} = (5+2)/2 = \frac{7}{2} = 3.5$

$\overline{u_7} = (4+3)/2 = \frac{7}{2} = 3.5$

$\overline{u_8} = (4+2)/2 = 3$

$\overline{u_9} = (5+4)/2 = \frac{9}{2} = 4.5$

$\overline{u_{10}} = (2+3)/2 = \frac{5}{2} = 2.5$

$\overline{u_{11}} = (4+1+5+2+2+4)/6 = 3$

$\overline{u_{12}} = (3+5)/2 = 4$

设 $t = \sqrt{1.5^2 + 1.5^2 + 0.5^2 + 0.5^2}$

再计算各用户与用户5的相似度，找出 2-最近邻，

$sim(1,5) = \dfrac{(2-\frac{4}{3})(2-3.5) + (1-\frac{4}{3})(3-3.5)}{\sqrt{(\frac{1}{3})^2 + (\frac{1}{3})^2 + (\frac{2}{3})^2} \quad t} = -0.456$

$sim(2,5) = \dfrac{(4-3)(2-3.5) + (2-3)(5-3.5)}{\sqrt{1^2+1^2} \quad t} = -0.949$

$sim(3,5) = \dfrac{(4-3.8)(5-3.5) + (4-3.8)(4-3.5) + (3-3.8)(3-3.5)}{\sqrt{0.8^2 + 1.2^2 + 0.2^2 + 0.2^2 + 0.8^2} \quad t} = 0.214$

$$\text{sim}(4,5) = \frac{(1-\frac{8}{3})(2-3.5) + (3-\frac{8}{3})(4-3.5)}{\sqrt{(\frac{4}{3})^2 + (\frac{5}{3})^2 + (\frac{1}{3})^2} \quad t} = 0.552$$

$$\text{sim}(6,5) = \frac{(2-3.5)(4-3.5)}{\sqrt{1.5^2 + 1.5^2} \quad t} = -0.158$$

$$\text{sim}(7,5) = \frac{(3-3.5)(2-3.5)}{\sqrt{0.5^2 + 0.5^2} \quad t} = 0.474$$

$$\text{sim}(8,5) = \frac{(4-3)(5-3.5) + (2-3)(3-3.5)}{\sqrt{1^2 + 1^2} \quad t \cdot} = 0.632$$

$$\text{sim}(9,5) = \frac{(4-4.5)(2-3.5)}{\sqrt{0.5^2 + 0.3^2} \quad t} = 0.474$$

$$\text{sim}(10,5) = \frac{(3-2.5)(2-3.5)}{\sqrt{0.5^2 + 0.5^2} \quad t} = -0.474$$

$$\text{sim}(11,5) = \frac{(5-3)(2-3.5) + (2-3)(5-3.5) + (2-3)(4-3.5) + (4-3)(3-3.5)}{\sqrt{1^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2} \quad t}$$

$$= -0.710$$

$$\text{sim}(12,5) = \frac{(5-4)(4-3.5)}{\sqrt{1^2 + 1^2} \quad t} = 0.158$$

相似度最高的 2个用户为 3 和9

故 $\text{pred}(5,1) = 3.5 + \dfrac{0.214(3-3.8) + 0.474(5-4.5)}{0.214 + 0.474}$

$$= 3.596$$

预测用户5对电影1的评分为3.596

比课件中给出的基于物品的评分结果(2.6)要高

2 问答题（言之有理即可）

2.1 社团推荐是社会网络分析中常见的问题。在仅考虑网络结构的情况下，已知网络中存在的若干社团，请结合课程内容中所介绍的推荐技术（包括链接推荐技术）与社团挖掘技术，设计一种给指定节点推荐社团的方法，并简述你的理由。

首先采用基于聚类的方法，用Q-Modularity指标确定合适的社团数，然后采用层次聚类或划分聚类的方法，将网络分为不同的社团，此时被指定的节点会被划分到某个社团中，则这个社团可以作为推荐。然后基于链接预测技术，预测该节点可能和哪些其他节点间存在边，并结合已有的边，推荐这些邻居（包括直接邻居和可能的邻居）所处的社团。这样，被推荐的社团都和该节点存在一定关联，推荐较为合理

2.2 试证明：在信息级联（Information Cascade）的定义下，信息传播最大化问题的目标函数具有"收益递减"特性，即给定两个集合 S、T 与集合外的节点 v，其中 S ⊆ T，满足：

$$f(S \cup \{v\}) - f(S) \quad \geq \quad f(T \cup \{v\}) - f(T)$$

∵ S ⊆ T   ∴ f(S) ≤ f(T)

对于集合外的节点 v

∴ $f(T \cap \{v\}) \geq f(S \cap \{v\})$

故 $f(T \cap \{v\}) + f(T) - f(S) \geq f(S \cap \{v\}) + f(T) - f(S)$

$f(T) - f(S) \geq [f(T) - f(T \cap \{v\})] - [f(S) - f(S \cap \{v\})]$

$f(T) - f(S) \geq [f(T) + f(\{v\}) - f(T \cap \{v\})] - [f(S) + f(\{v\}) - f(S \cap \{v\})]$

∵ $f(T \cup \{v\}) = f(T) + f(\{v\}) - f(T \cap \{v\})$
  $f(S \cup \{v\}) = f(S) + f(\{v\}) - f(S \cap \{v\})$

∴ $f(T) - f(S) \geq f(T \cup \{v\}) - f(S \cup \{v\})$

∴ $f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$