基本思想：用户会参加拥有其感兴趣的主题的社团举办的活动，即用户-主题-社团-活动；

同时用户会受到三类社交约束的限制：

1） 拥有相同兴趣的用户会参加类似的活动（标签相似性）；

2） 同一个社团的成员可能参加相同的活动；

3） 参加过同一个活动的成员可能参加相同的活动；

算法设计：

1） 整体思路：

- 定义 user-event 矩阵 UE，矩阵行为全体 user，矩阵列为训练集中的 event，每个元素为用户是否参加过此 event，如果参加过，则元素为 1，未参加过，元素为 0。

例如用户 i 参加过 event j，则 (i,j)=1，反之, (i,j)=0;

由训练集得到的 UE 矩阵可以记为 UE_train；

有测试集得到的 UE 矩阵记为 UE_test；

两者的行都为所有用户，但是列的选取，即 event 的选取则根据各自数据集中的 event 来定义；

- 定义特征矩阵 GT, 行为社团 group，列为主题 Topic, 元素定义如 UE, 也是 0，1 元素构成的稀疏矩阵；
- 定义特征矩阵 UT，行为用户 user，列为主题 topic，元素定义如 UE, 也是 0，1 元素构成的稀疏矩阵；

- 定义关联矩阵 GE，行为社团 group，列为活动 event（此时的 event 为训练集+测试集中的 event），元素定义如 UE，也是 0，1 元素构成的稀疏矩阵;
- 算法的输入：UE_train，GT, UT, GE;
- 算法的输出：UE_predict;
- 算法结果的检验：UE_predict 和 UE_test 之间的比较，即按照如下矩阵依次核对 UE_test 中的每个元素(i,j)，记录 TP, FP, TN, FN 的数量：

	UE_test 中为 1	UE_test 中为 0
UE_predict 中为 1	TP	FP
UE_predict 中为 0	FN	TN

按照统计出的 TP，FP，FN 和 TN 的数量，计算下面四个指标：

Precision= TP/(TP+FP);

Recall= TP/(TP+FN);

Accuracy= (TP+TN)/(TP+TN+FP+FN)

F1 score=2*Precision*Recall/(Precision+Recall)

2) 方案 A：兴趣/主题网络的影响

a) 根据 UT 构建用户之间的主题相似性网络 UTS, 其实现也是一个矩阵，其行、列均为用户，每个元素代表用户之间的主题相似性，即两者相同的主题个数与两者总主题个数之比。例如用户 i 和用户 j 之间相同的主题为 2 个，则(i,j)=2/(i 的主题个数+j 的主题个数);

b) 然后根据 UE_train 计算两个用户 a 和 b 的“活动相似度”如下：

$$Sim(a,b)=UTS(a,b)+\frac{\sum_{e\in event(E\_train)}\left(r_{a,e}-\overline{r}_a\right)\left(r_{b,e}-\overline{r}_b\right)}{\sqrt{\sum_{e\in event(E\_train)}\left(r_{a,e}-\overline{r}_a\right)^2}\sqrt{\sum_{e\in event(E\_train)}\left(r_{b,e}-\overline{r}_b\right)^2}}$$

其中 event(E_train)为 UE_train 中全部 event 的集合，|event(E_train)|代表训练集中所有 event 的数量；

UTS(a,b)代表 UTS 矩阵中(a,b)的元素值，即用户 a 与 b 的主题相似性；

$r_{a,e}$ 代表用户 a 是否参加过活动 e，即矩阵 UE_train 中(a,e)的值；

$\overline{r}_a$ 代表用户 a 在 UE_train 全部 event 参与情况的平均值，其等于：

$$\overline{r}_a=\frac{\sum_{e\in event(E\_train)}r_{a,e}}{\left|event(E\_train)\right|}$$

可以将每对用户之间的活动相似性记录下来存成一个矩阵 Sim，避免重复计算；

c) 接下来每个用户根据计算得到的活动相似度选取最接近（相似性最高）的前 K 个用户（K 是超参数），构建其邻居集合 Neighbor. 例如用户 a 的邻居集合记为 Neighbor(a)；

d) 对每一个用户 a，遍历 UE_test 中的每个活动 e，并根据 GE，GT 和 UT 找到 e 所对应的主题 Te（活动-社团-主题），计算预测值如下：

$$pred(a,e)=\overline{r}_a+\frac{\sum_{b\in Neighbor(a)}sim(a,b)\cdot I\left(T_e\in NT_a\right)}{\sum_{b\in Neighbor(a)}sim(a,b)}$$

其中 NT_a 表示用户 a 的邻居（包括其自身）的主题的并集

$I\left(T_e\in NT_a\right)$ 为示性函数：

$$I\left(T_e \in NT_a\right)=\begin{cases}1, & T_e \in NT_a \\ 0, & T_e \notin NT_a\end{cases}$$

e) 如果预测值 pred(a,e)大于某个阈值θ（超参数，例如 0.5），则预测 a 接受 e，即 1，反之为 0;

总结: 这里的社交约束本质上是扩大了一个用户 a 的主题范围从而增加了选择某个活动的可能性;

注：测试集中不在 yes/no 中的用户也视为标签为 0（即 0 标签人物=no 的用户+不在 yes 和 no 中的用户）