第一题（Mathematical Analysis of Algorithm）(20 分)

（1） 论文中的In Situ Permutation算法最坏情形的时间复杂度是多少？对应的输入是怎样的？（5分）

（2） 如果按论文中所述增加了tally变量，其最坏情形的时间复杂度是怎样的？对应的输入又是怎样的？（5分）

（3） 在增加了tally变量后，第6行代码循环次数平均减少多少次？（注：论文中的结论有误）（5分）

（4） 如果再给定*p(j)*的反函数*p-1(j)*，那么是否可以对算法进一步优化？最坏时间复杂度和平均复杂度有怎样的变化？（5分）

第二题（Primes is in P）（20分）

（1） 文章的引言部分回顾了一些重要的素性测试算法。请判断下表第一列中的素性测试算法是否满足表中第一行的性质。若满足，请在表中对应位置打"√"；若不满足，请打"X"；若无法判断，请写"—"。（6分）

	unconditional	deterministic	polynomial-time
古希腊，Sieve of Eratosthenes			
1974年，Solovay and Strassen在文[SS]中给出的算法			
1975年，Miller在文[Mil]中给出的算法			
1983年，Adleman, Pomerance, Rumely在文[APR]中给出的算法			

（2） 请阅读文章引言第二段前7行，即"Let PRIMES denote ...if n is a prime number"，解释文中的复杂度$\Omega(\sqrt{n})$为什么是效率低的(inefficient)。（3分）

（3） 请问：PRIMES问题是NP问题吗？是NP难问题吗？是NP完全问题吗？（只给出判断"是"或"否"，无需说明理由。）（3分）

（4） 文章在第3节，介绍了AKS算法中用到的一个重要术语"a模r的阶"及符号$o_r(a)$。请设计一个算法，对于给定的且满足$(a,r) = 1$的整数a和正整数r，求出$o_r(a)$.（5分）

（5） Wikipedia里关于AKS Primality Test，有如下评论：*"While the algorithm is of immense theoretical importance, it is not used in practice."*请谈谈你对这句话的理解。（3分）

第三题（LINE: Large-scale Information Network Embedding）（20分）

（1） 简述Network中二阶相似性的定义，至少举一个生活中的例子。(5分)

（2） 论文中，(5)式和(6)式在模型优化的意义上等价，请写出(5)到(6)的数学推导过程。(5分)

$$O_2 = \sum_{i \in V} \lambda_i d(\hat{p}_2(\cdot|v_i), p_2(\cdot|v_i)), \tag{5}$$

$$O_2 = -\sum_{(i,j) \in E} w_{ij} \log p_2(v_j|v_i). \tag{6}$$

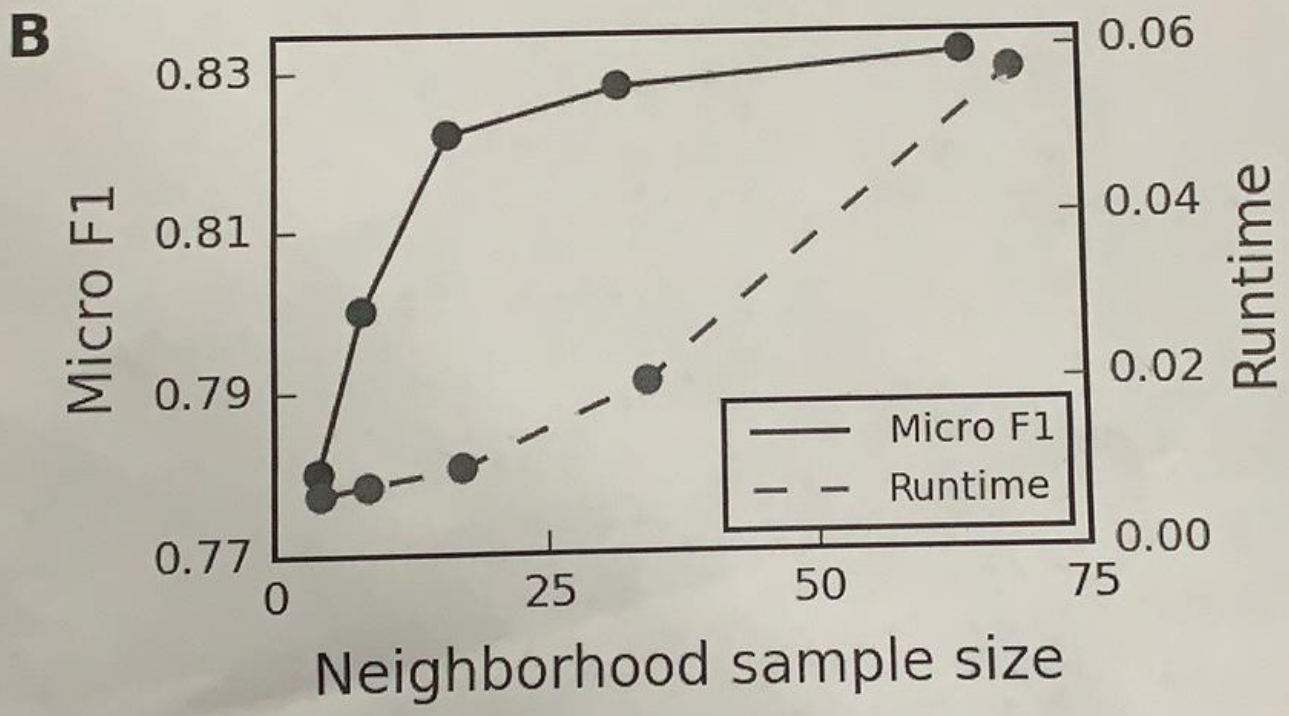
（3） 为什么要在优化过程中进行负采样(Negative Sampling)？简述其思路。(5分)

（4） 请发挥想象，类比LINE的思路，提出一个更高阶相似性(比如3rd Order Proximity)的定义及损失函数。(5分)

第四题（Inductive Representation Learning on Large Graphs）（20分）

（1） 简述GraphSAGE模型的思路，以及四种邻域聚集器(Aggregator)的思想。(8分)

（2） 请说明下图中横纵坐标及两条曲线的意义。两条曲线的走势说明了什么现象？GraphSAGE采用的哪一种技术细节是有效的？(6分)



（3） 以GraphSAGE等为代表的邻域聚集型网络表示方法有何优点和局限？（可对比LINE或者其他你知道的方法，言之成理即可）(6分)

$\sum_{i \in V} \cdot d_i \, d\left(\frac{w_i}{d_i}\right)$