

平成 20 年度 夏入試

東京大学情報理工学系研究科創造情報学専攻

プログラミング

注意事項

1. 試験開始の合図まで，この問題冊子を開いてはいけません。
2. この表紙の下部にある受験番号欄に受験番号を記入して下さい。
3. 答案用紙と下書き用紙が 1 枚ずつ配られる。それぞれに受験番号を記入して下さい。途中で余白が足りなくなったら，試験監督に申し出下さい。すべての問題の解答は答案用紙に書き下さい。Q2 以降の問題は，コンピュータの出力を参照して，答案用紙に解答して下さい。
4. 各受験者に配られた USB メモリに ASCII コードで書かれたファイル `edges.txt` が 1 個入っている。改行はすべて CR と LF の対で書かれている (CR は carriage return，LF は line feed である)。

試験開始前に，USB メモリから上記のファイルを自分の PC にコピーして下さい。各ファイルの中身を見て，ASCII 文字で書かれた自然数の対がたくさんに行にわたって入っていることを確認し，PC から手を離して下さい。ファイルにアクセスできない，あるいは中身が ASCII 文字列として読めないなどの場合は試験監督に申し出下さい。言うまでもなく，USB メモリの中身は全受験者に共通である。
5. プログラミング言語は各自の得意なものを使用して下さい。
6. プログラミング言語のマニュアルは 1 冊に限り試験中に参照してもよい。
7. 試験終了時まで，自分の PC 上に受験番号名のディレクトリ/フォルダを作成し，作成したプログラムおよび関連ファイルをその下にコピーして下さい。作成したディレクトリ/フォルダを各受験者に渡された USB メモリにコピーして下さい。
8. 試験終了時に，USB メモリ，答案用紙，下書き用紙を回収する。
9. 回収後，試験監督が周回し，各受験者の結果をごく簡単に確認するので，そのまま座席で待機して下さい。全員の確認が終わるまで部屋を出てはいけません。
10. 午後のプログラミングの口頭試問中にプログラムの動作をより精密に確認する。各自の PC 上でプログラムがなるべくすぐに実行できるようにしておきなさい。
11. 全員の確認が終了した後，各自の PC とこの問題冊子を残し，部屋から退出して下さい。

受験番号 _____

人々の関係、情報流通、交通網など、社会構造を分析するときには、グラフでのモデル化がよく行なわれる。例えば、航空路線グラフでは、空港を頂点で表し、直通便がある空港同士を頂点を結ぶ辺で表す。このような社会構造を表すグラフの解析に便利ないろいろな特徴量を計算しよう。

n 個の頂点 (1 から n までの番号を振る) をもつすべての可能なグラフの集合を $G(n)$ と書くことにする。頂点 v_i と頂点 v_j を結ぶ辺を (v_i, v_j) と書く。以下で考慮するグラフは、同一頂点を結ぶ円弧のような辺がなく、同じ頂点对を結ぶ 2 本以上の辺がない単純なグラフとする。つまり、航空路線グラフの場合、航空会社が違う同一路線は 1 本で表す。

n 個の頂点をもつグラフを文字列で表すのに、辺 (v_i, v_j) を i と j を空白で区切った頂点对で表し、それらをさらに空白または改行で区切って並べて書く。Q2 以降、 $n = 100$ と固定し、 N は変えるが同じ頂点对の並びを利用する。その場合、 N 番目以降の頂点对は無視する。

Q1 以下の頂点对の列で表される $G_1 \in G(9)$ を答案用紙に描きなさい。辺は交差しないようにし、頂点は次ページの図 1 にならって、 の中に頂点番号を入れて書きなさい。

3 1 4 1 5 9 2 6 5 3 5 8 9 7 9 3 2 3 8 4

Q2 グラフに関する以下の定義を読み、それに続く問題に答えなさい。以下に定義された概念を処理するのに無駄がなく、かつ便利なデータ構造を考えることが重要である。

- (1) 頂点 v の次数とは、 v から何本の辺が出ているかを表す非負整数である。航空路線グラフでは空港から何本の路線が出ているかを示す。
- (2) 頂点 v のクラスタ係数とは、 v と直接結ばれている隣接頂点の集合 $W = \{w_1, \dots, w_k\}$ の中の頂点同士の間の辺の数を、可能な最大数 ${}_kC_2 = k(k-1)/2$ で割った実数である。ただし、次数 1 以下の頂点のクラスタ係数は 0 とする。図 1 にクラスタ係数の例を示す。航空路線グラフでは、クラスタ係数の高い空港は、その周囲の空港間に緊密な路線網があることを示す。
- (3) グラフの平均クラスタ係数とは、そのグラフのすべての頂点のクラスタ係数の算術平均である。航空路線グラフでは、航空網全体にどれくらいの割合で直通便の 3 角形があるかを示す特徴量である。
- (4) 2 つの頂点 v_i, v_j を結ぶ路とは v_i, v_j を結ぶ辺の列、 $(w_0, w_1), (w_1, w_2), \dots, (w_{k-1}, w_k)$ 、ただし、 $w_0 = v_i, w_k = v_j$ である。 k を路の長さと呼ぶ。また、2 つの頂点 v_i, v_j の距離とは、 v_i と v_j を結ぶ最短の路の長さのことである。ただし、 $v_i = v_j$ のときは距離を 0、 v_i と v_j を結ぶ路が存在しないときは距離を ∞ とする。航空路線グラフでは、距離が短ければ、目的地までの乗り換え回数が少ない。
- (5) グラフの連結成分とは、お互いを結ぶ路がある頂点のすべてとそれらを結ぶ辺をすべて含む部分グラフである。連結グラフとは連結成分が 1 個しかないグラフである。
- (6) 連結グラフ G の直径とは、 G の 2 つの頂点の距離の最大値である。また、平均直径とは G の異なる 2 つの頂点の距離の算術平均である。航空路線グラフの平均直径が小さいと、平均的には便利な航空路線が用意されていると言える。

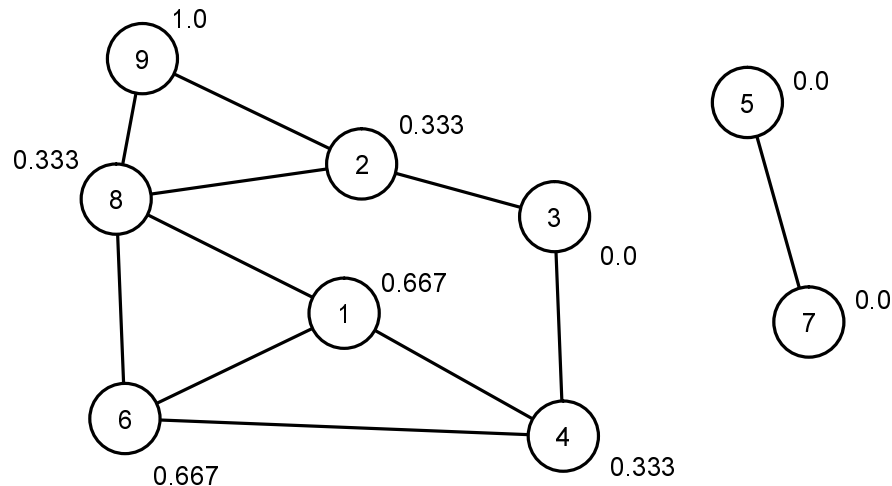


図 1. グラフの図示とクラスタ係数 (頂点の傍に書いてあるのがクラスタ係数)

ファイル `edges.txt` に書かれた頂点对の並びの最初の 181 個を読み, グラフ $G_2 \in \mathbf{G}(100)$ をメモリ内に作成しなさい. ファイル `edges.txt` には, 100 個の頂点間を結ぶすべての可能な 4,950 個の頂点对が 1 行に 1 個ずつ書かれている. 以下のすべての問題は, コンピュータの出力を参照して, 答案用紙に解答しなさい. 複数の数が並ぶときには間に空白や改行を適宜入れて書きなさい. 小数点のある数は, 末尾の 0 を除いて少なくとも小数点以下 3 桁まで表示すること.

Q2-1 G_2 のそれぞれの連結成分の頂点の個数を降順に出力しなさい.

Q2-2 G_2 の頂点 1 ~ 10 のクラスタ係数をこの順にすべて出力しなさい.

Q2-3 G_2 の平均クラスタ係数を出力しなさい.

Q2-4 ファイル `edges.txt` の頂点对を使うとき, 辺の個数 N を最低いくつにすれば, グラフが連結グラフになるかを出力しなさい. このときにできたグラフ G_3 の平均クラスタ係数を出力しなさい.

Q2-5 ファイル `edges.txt` の残りの頂点对を使い, グラフ G_3 から辺をさらに 100 個増やしたグラフ G_4 での平均クラスタ係数を出力しなさい. なお, G_2 は意図的に作成したグラフなので, G_4 の平均クラスタ係数は G_3 に比べて減少する.

Q3 グラフ G_3, G_4 における頂点 27 と頂点 63 の距離をそれぞれ出力しなさい. また, これらのグラフを仮想的な航空路線図として見たとき, G_4 が G_3 に比べてどれくらい便利になっているかを確認するために, G_3 と G_4 の平均直径をそれぞれ出力しなさい.

Q4 初めて連結グラフになる G_3 において, 直径が有限の値になるが, その後, `edges.txt` の残りの頂点对に従って辺を増やしていったとき, 辺の数がいくつになるたびに直径が減るかを, 辺の数と直径の対として順次出力しなさい. 最後の出力は直径が 1 になる 4950 1 となる. なお, G_3 の辺の個数と直径を最初に出力する必要はない.

2008 Summer Entrance Examination

Department of Creative Informatics
Graduate School of Information Science and Technology
The University of Tokyo

Programming

INSTRUCTIONS

1. Do not open this problem brochure until the signal to begin is given.
2. Write your examinee ID below on this cover.
3. An answer sheet and a draft sheet are delivered. Write down your examinee ID on both sheets. If you need another sheet in the examination, raise your hand and request it to the test proctor. **Write your answers to all the questions on the answer sheet. Write your answers referring to your computer output for the questions from Q2.**
4. The USB memory delivered beforehand to each examinee contains an ASCII text file: `edges.txt`. Newline is represented by a pair of carriage return (CR) and linefeed (LF) in this file.
Before examination starts, copy this file to your PC and browse the file. Make sure you can see a lot of number pairs each in a line, and keep your hands away from your PC. If you cannot read the file properly, consult the test proctor. Needless to say, the contents of the USB memory is common to all examinees.
5. You may choose your favorite programming language.
6. You may consult only one printed manual of the programming language in the examination.
7. By the end of the examination, make a directory/folder on your PC, whose name is the same as your examinee ID, and put your program files and related files under the directory/folder. Copy the directory/folder into the USB memory.
8. At the end of the examination, the USB memory, the answer sheet and draft sheet are collected.
9. After the collection, stay at your seat, until all examinee results have been checked briefly by the test proctor.
10. After the brief check, try to retain your program execution environment on the PC so as to be able to resume it as soon as possible at the oral examination.
11. **Leave your PC and this brochure together in the room** for the oral examination and leave the room until you are called.

Examinee ID _____

Social structures such as human relation, information circulation, and traffic network are often modeled by graphs. For example, in an airline network graph, airports are denoted by nodes, and direct flights between airports are denoted by edges connecting the nodes. Let us calculate various metrics of graphs, which may be useful to analyze social structures.

The set of all possible graphs consisting of n nodes (numbered from 1 to n) is denoted $\mathbf{G}(n)$. The edge between nodes v_i and v_j is denoted (v_i, v_j) . We will consider only simple graphs where there is no edge between the same node and no duplicate edges between the same pair of nodes. That is, in the case of the airline network graph, only one edge is spanned between two airports even if different airline companies have direct flights.

Let us denote a graph consisting of n nodes in a plain text as a sequence of the node pairs separated by space(s) or newline, where a node pair i and j separated by a space represents an edge (v_i, v_j) . Since we will fix $n = 100$ from Q2 and use the same node pair sequence for varying N , let us ignore the node pairs after N -th pair.

Q1 Draw the graph $G_1 \in \mathbf{G}(9)$ denoted in the following node pair sequence on the answer sheet. Circle each node number as illustrated in Figure 1, and do not cross edges.

3 1 4 1 5 9 2 6 5 3 5 8 9 7 9 3 2 3 8 4

Q2 Read the following definitions about graphs, and answer the questions. It is important to design efficient and convenient data structures for the questions below.

- (1) The **degree** of a node is the number of edges connected to the node. In the airline network graph, it shows the number of available direct flight lines from the airport.
- (2) The **cluster coefficient** of a node v is the number of edges among the node set $W = \{w_1, \dots, w_k\}$ of v 's adjacent (directly connected) nodes divided by its maximum possible number ${}_kC_2 = k(k-1)/2$. For nodes whose degree is less than or equal to 1, the cluster coefficient is defined 0. The cluster coefficient is illustrated in Figure 1. In the airline network graph, a dense airline network can be observed in the vicinity of an airport with higher cluster coefficient.
- (3) The **average cluster coefficient** of a graph is an arithmetic means of the cluster coefficients of all the nodes. In the airline network graph, it shows how much triangular direct flight networks exist in proportion to the total network.
- (4) The **path** between nodes v_i and v_j is a sequence of edges $(w_0, w_1), (w_1, w_2), \dots, (w_{k-1}, w_k)$, where $w_0 = v_i, w_k = v_j$. k is called the length of the path. The **distance** of nodes v_i and v_j is the length of the shortest path between v_i and v_j . Let the distance of the identical nodes be 0, and the distance of nodes be ∞ if there is no path between them. The less the distance is, the fewer connections are needed for the destination in the airline network graph.
- (5) The **connected component** of a graph is a subgraph consisting of all nodes connected by paths, and edges that connect the nodes. A **connected graph** is a graph which has only one connected component.
- (6) The **diameter** of a connected graph G is a maximum distance of all possible two nodes of G . The **average diameter** is the arithmetic means of all distances of mutually different nodes of G . Smaller average diameter of the airline network graph may suggest more convenient airline network.

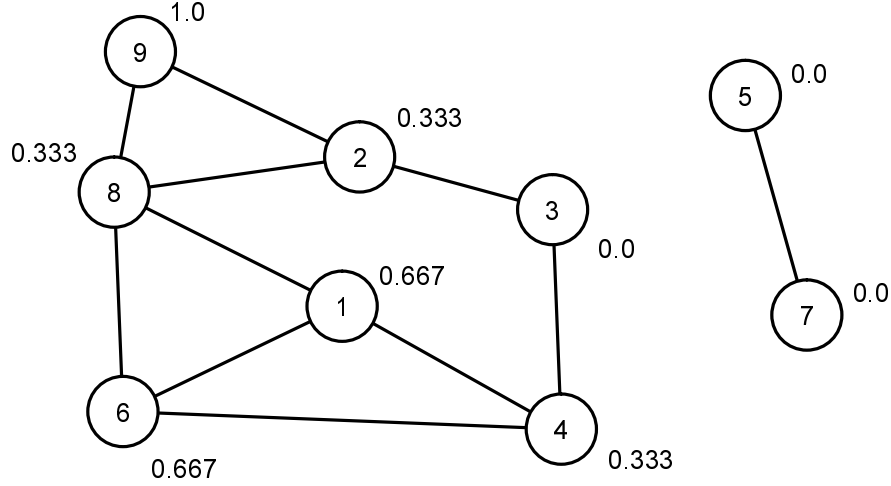


Fig 1. Graph, and Cluster Coefficients attached to the nodes

Read the first 181 node pairs written in the file `edges.txt`, and represent a graph $G_2 \in \mathbf{G}(100)$ in your computer memory. The file `edges.txt` contains all possible 4,950 node pairs or edges for 100 nodes, each in a line. **For the subsequent problems, you should write your answers on the answer sheet, referring to your computer output.** More than one numbers should be separated appropriately by space or newline. Decimal point numbers should be shown up to at least three places of decimals other than trailing zeros.

Q2-1 Output the numbers of the nodes in all the connected components of G_2 in descending order.

Q2-2 Output the cluster coefficients of nodes 1 to 10 in this order.

Q2-3 Output the average cluster coefficient of G_2 .

Q2-4 By using the node pairs in `edges.txt`, at least what N will make the graph connected? Output the N , and output the average cluster coefficient of the obtained graph G_3 .

Q2-5 Output the average cluster coefficient of the graph G_4 obtained by using 100 more node pairs in `edges.txt` after obtaining G_3 . Note that the average cluster coefficient of G_4 is less than that of G_3 , because G_3 is created somehow intentionally.

Q3 Output the distances of node 27 and node 63 in G_3 and G_4 , respectively. Then output the average diameters of G_3 and G_4 , respectively, to estimate the average usefulness of the airline network, assuming these graphs model airline networks.

Q4 As N increases, the diameter of the first connected graph G_3 becomes finite for the first time. If we add edges by using remaining node pairs in the file `edges.txt`, the radius of the obtained graph will decrease occasionally. Output all the pairs of the number of edges and the corresponding radius one by one, when an added edge decreases the radius. The last output should be 4950 1, where the last added edge decreases the radius to 1. It is not needed to output the number of edges of G_3 and its radius at first.