MorphoGraphs and Imagery - Exercises #1 :

Cyber-profiler

Originally proposed by Michel Couprie

In a software helping to solve criminal cases, an undirected graph is used to find correlations between different cases. A vertex of this graph represents an event, a person or a place, and two vertices share a common edge if a link between them was established, for example, if two people p_1 and p_2 stayed in a same location, or if they are in a family or business relationship, or if two events E_1 and E_2 involve a same person...

This software is used to automatically find out groups of vertices that are linked by many edges. Such coincidences may indicate links between several past or present cases. In the following, our goal is to formalize and quantify this notion of "coincidence" and to propose a method to detect them.

Let $G = (E, \overrightarrow{\Gamma})$ be a graph and let X be a subset of E. The *subgraph of G induced by X* is the graph $G_X = (X, \overrightarrow{\Gamma} \cap [X \times X])$. In other words, an arc of G is also an arc in G_X if (and only if) its two extremities are in X.

We say that two vertices x and y are *adjacent in G* if the unordered pair $\{x, y\}$ is an edge of G (or more precisely an edge of the undirected graph $(E, \overline{\Gamma})$ associated with G).

In the following, we assume that $G = (E, \overrightarrow{\Gamma})$ is a symmetric graph without loop. We denote by n and m the numbers of vertices and of edges of G ($n = |E|$ and $m = |\overrightarrow{\Gamma}|$).

Let X be a subset of E, and let k be a positive integer. We say that X is *k-linked (for G)* if any vertex $x \in X$ is adjacent to at least k vertices in the subgraph of G induced by X.

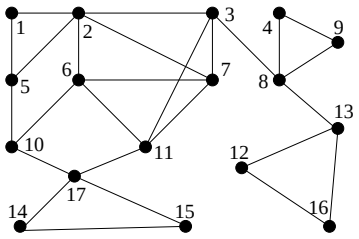

Figure 1

a) In the graph represented in Figure 1, is there
- a subset of 3 vertices that is 1-linked but not 2-linked ?
- a subset of 4 vertices that is 2-linked but not 3-linked ?
- a subset of 5 vertices that is 3-linked ?
- a subset of vertices that is 4-linked ?

For each positive answer, give the corresponding set.

b) Let $x \in E$. We denote by $d(x, G)$ the degree of x in the graph G, that is the number of vertices

of G that are adjacent to x in G. Give a compact characterization (*i.e.,* another definition) of a set that is k-linked using the above notation.

c) Let $X \subseteq E$. What do you think of the following statement :

$$X \text{ is 1-linked if and only if } X \text{ is connected}^1.$$

Justify your answer.

In our application to search for links between criminal cases, we look for (nonempty) sets that are k-linked, where k is maximal, the number k measuring the effectiveness of a cluster of coincidences. Is it always enough to show the connection between two cases ? The next item provides an answer to this question.

d) Let $X \subseteq E$ with $|X| = p \geq 2$. Prove that if X is k-linked with $k \geq p/2$, then X is connected.

e) Let $k \in \mathbb{N}$. Propose a method to automatically find a subset X of E that is k-linked if such subset exists. In the case of non-existence of such subset, your method must return a negative answer. You can describe your method in English. Further items will ask for an algorithmic scheme.

f) Propose a detailed algorithmic scheme for the method described in e). The proposed algorithm does not need to be optimal, a further item will ask for an optimal time-complexity algorithm.

g) Assess the time complexity of your algorithm.

h) If your algorithm (questions f-g) does not run in linear-time, propose another algorithm whose time-complexity is $O(n + m)$. Justify the complexity of this algorithm.

i) Run ("by hand", on your sheet of paper) the algorithm on the graph of Figure 1, for $k = 3$

We are now going to look for arguments that justify the correctness of the proposed method. In other words, we are going to prove that the method always produces the expected results (according to the definition).

Let X be a subset of E. We say that X is a *clique* for G if any two vertices of X are adjacent to each other. A clique X for G is said to be *maximal (with respect to inclusion)* whenever there is no clique that strictly includes X.

j) What is the link between cliques and k-linked subsets ? Illustrate the notion of a maximal clique by some examples. Give an example of a graph that has two disjoint maximal cliques.

k) Let $k \in \mathbb{N}$. Let X be a subset of E that is k-linked and that is maximal for the inclusion relation. Prove that this subset is unique.

l) Let X be a subset of E and let Y be a subset of E that includes X. Prove that X is k-linked for G if and only if X is k-linked for the subgraph G_Y of G induced by Y.

m) Let $k \in \mathbb{N}$. We denote by E_k the subset of E that is k-linked and that is maximal for the inclusion relation (see item k). Prove that the result of the method found at item e) is precisely E_k.

n) Prove that for any two integers k and k' such that $k' \geq k$, we have $E_{k'} \subseteq E_k$.

o) Propose a method to automatically compute the largest integer k such that there exists a nonempty subset X of E that is k-linked. Redundant computations should be avoided based on properties e) and n).

[1] A subset X of E is connected if the subgraph G_X of G induced by X has only one connected component.