Un problème de digicode

Comme on peut le constater lorsque l'on se trouve face à un digicode, l'appareil fonctionne (la plupart du temps) de la façon suivante : l'utilisateur compose une suite de lettres (ou de chiffres) afin de composer un mot qui soit le "sésame" de la porte concernée. Supposons par exemple que le mot recherché soit composé de cinq caractères. L'appareil va, à chaque nouveau caractère choisi par l'utilisateur, tester le mot constitué par les cinq derniers caractères tapés par la personne.

Ainsi, si l'utilisateur rentre la séquence : 123456, la machine testera les mots 12345 et 23456.

Le problème qui nous intéresse ici est de savoir quel est le nombre minimal de caractères qu'il faut taper afin que la machine teste tous les mots possibles, et qu'ainsi la porte s'ouvre à coup sûr. Formellement :

Problème 1 *Soient n et p deux entiers.*

Soit un alphabet à p lettres. Quelle est la taille du plus petit mot contenant tous les mots de n lettres sur cet alphabet ?

Remarquons tout d'abord qu'il y a p^n mots de n lettres sur un alphabet de p lettres. Nous appelerons "dictionnaire" l'ensemble de ces mots. On peut en déduire deux bornes pour la taille du mot minimal répondant au problème :

- Si l'on met bout à bout tous les mots de notre dictionnaire, le mot obtenu contient bien sûr tous les mots possibles. Le mot obtenu est de taille $n \times p^n$. En effet on met bout à bout p^n mots de n lettres chacun.
  Naturellement, ce mot est loin d'être optimal.
- Inversement, l'hypothèse optimiste est qu'il existe un mot sans redondance, c'est-à-dire qui contienne une et une seule fois chaque mot de n lettres. Un tel mot répondrait à notre problème, puisque l'on ne peut pas faire plus court.
  Sa taille serait p^n+n-1. En effet la séquence obtenue "épuiserait" à chaque nouvelle lettre tapée un mot de notre dictionnaire. Mais ce à partir de la n-ième lettre tapée, naturellement.

Nous allons de fait montrer que cette dernière hypothèse est la bonne. Pour cela, il nous faut quelques outils de théorie des graphes. En quelques mots, on construit un graphe orienté dont les sommets sont les mots de $n-1$ lettres, et les arcs les mots de n lettres. Il s'agit de trouver dans ce graphe un chemin qui parcours tous les arcs une et une seule fois.

L'objet de ce qui suit est de trouver une condition suffisante sur un graphe pour qu'un tel chemin existe. Nous verrons ensuite que le graphe obtenu dans ce problème précis vérifie bien cette condition.

1 Un résultat sur les graphes orientés

On peut trouver les résultats démontrés dans cette partie dans le chapitre 12 de [1], nous avons ici essayé d'isoler le plus possible le résultat précis qui permet de résoudre notre problème de digicodes. On trouvera également une introduction à la théorie des graphes claire, complète et accessible dans [2].

Il est cependant nécessaire d'utiliser le vocabulaire de la théorie des graphes, aussi nous allons être amenés à définir un certain nombre de notions. Malgré le caractère parfois aride de ces définitions, celles-ci correspondent à des idées assez simples, que nous essaierons de reformuler de façon intuitive.

1.1 Quelques définitions

Définition Un graphe orienté G est la donnée d'une paire $(\mathrm{S}(\mathrm{G}), \mathrm{A}(\mathrm{G}))$, où $\mathrm{S}(\mathrm{G})$ est un ensemble fini dont les éléments sont appelés sommets de G et $\mathrm{A}(\mathrm{G})$ est un ensemble fini de couples ordonnés (distincts) d'éléments de $\mathrm{S}(\mathrm{G})$, appelés arcs de G. Un arc $(\mathrm{S}_1, \mathrm{S}_2)$ sera aussi noté $\mathrm{S}_1\mathrm{S}_2$, et l'on dira que cet arc relie S_1 à S_2. S_1 est alors appelé l'origine de l'arc et S_2 son extrémité.

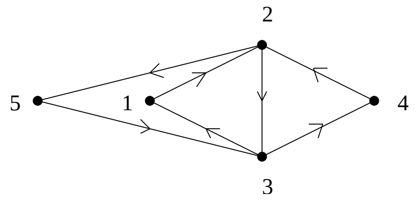
Intuitivement, un graphe orienté n'est rien d'autre qu'un ensemble de points reliés par des flèches.

Définition

- **(i)** Un chemin de longueur r du graphe orienté G est une suite $(\mathrm{S}_0, \ldots, \mathrm{S}_r)$ de sommets telle que pour tout i dans $[\![0 ; r-1]\!]$ il existe un arc reliant S_i à S_{i+1}. Ces r arcs sont de plus supposés deux à deux distinct. S_0 est appelé origine du chemin, et S_r extrémité du chemin.
- **(ii)** Le chemin $(\mathrm{S}_0, \ldots, \mathrm{S}_r)$ est appelé un circuit (ou chemin fermé) s'il vérifie $\mathrm{S}_0 = \mathrm{S}_r$.
- **(iii)** Un circuit élémentaire est un circuit dont tous les sommets sont distincts (mis à part le premier et le dernier bien sûr).

Exemple :

Dans le graphe suivant, $(1,2,3,1)$ et $(1,2,3,4,2,5,3,1)$ sont deux circuits possibles, mais bien sûr seul $(1,2,3,1)$ est un circuit élémentaire. Le circuit $(1,2,3,4,2,5,3,1)$ passe deux fois par le sommet 2 (et le sommet 3).



Propriété 1.1 *Tout circuit peut se décomposer en circuits élémentaires.*

En d'autres termes, étant donné un circuit C, on peut trouver un ensemble de circuits élémentaires $\mathrm{E}_1, \ldots, \mathrm{E}_p$ de façon à ce que l'ensembles des arcs de C soit l'union des ensembles des arcs de E_1, $\mathrm{E}_2, \ldots, \mathrm{E}_p$.

Ainsi le circuit $(1,2,3,4,2,5,3,1)$ de l'exemple précédent se décompose en deux circuits élémentaires : les circuits $(1,2,3,1)$ et $(2,5,3,4,2)$.

Nous allons démontrer ce résultat par récurrence sur la longueur du circuit C considéré.

- Si le circuit C est de longueur 1, il est de la forme S, S, et n'est constitué que d'un seul arc SS. Il est alors bien sûr élémentaire.
- Supposons le résultat vrai jusqu'à l'ordre n. Soit alors $\mathrm{C} = (\mathrm{S}_i)_{0 \leq i \leq n+1}$ un circuit de longueur $n + 1$.
  Soit C est lui-même un circuit élémentaire, auquel cas il n'y a rien à démontrer, soit C n'est pas élémentaire. C'est-à-dire qu'il existe une paire d'indices $i < j$, vérifiant $(i, j) \neq (1, n + 1)$, telle que $\mathrm{S}_i = \mathrm{S}_j$.
  Nous pouvons alors décomposer le circuit C en deux circuits $\mathrm{S}_i, \ldots, \mathrm{S}_j$ et $\mathrm{S}_0, \ldots, \mathrm{S}_i, \mathrm{S}_{j+1}, \ldots, \mathrm{S}_{n+1}$, qui sont eux-mêmes des circuits strictement moins longs, donc de longueur au plus n. Par hypothèse de récurrence, ils peuvent eux-même être décomposés en circuits élémentaires, et donc C également.

Par le principe de récurrence, on peut donc conclure que tout circuit d'un graphe orienté se décompose en circuits élémentaires.

□

1.2 Décomposition en circuits disjoints

Définition

(i) La valence sortante $d_+(\mathrm{S})$ d'un sommet S est le nombre d'arc partant de ce sommet.

(ii) La valence entrante $d_-(\mathrm{S})$ d'un sommet S est le nombre d'arc arrivant à ce sommet.

Définition Un graphe orienté G est dit pseudo-symétrique si tous ses sommets S vérifient $d_-(\mathrm{S}) = d_+(\mathrm{S})$. En d'autres termes, G est pseudo-symétrique, si pour tout sommet de G, il y a le même nombre d'arcs qui arrivent à ce sommet que d'arcs qui en partent.

Propriété 1.2 *Un graphe orienté* G *est pseudo-symétrique si et seulement si il est réunion de circuits élémentaires disjoints.*

L'un des sens est immédiat. Si G est réunion de circuits élémentaires disjoints, alors pour tout sommet S de G, les valences sortantes et entrantes de S sont égales au nombres de circuits élémentaires passant par S, donc égales entre elles. Et le graphe orienté G est donc pseudo-symétrique.

Réciproquement, nous allons montrer par récurrence sur le nombre d'arcs que tout graphe pseudo-symétrique est réunion de circuits élémentaires.

- Si G ne possède aucun arc, le résultat est immédiat.
- Supposons le résultat vrai pour tout graphe orienté pseudo-symétrique d'au plus n arcs. Soit G un graphe orienté possédant $n + 1$ arcs.
  Tout d'abord, nous allons montrer par l'absurde que G contient un circuit, donc un circuit élémentaire. Supposons donc que G n'ait pas de circuit.
  Partant d'un sommet S_1 quelconque de G dont la valence sortante est non nulle (rappelons que G possède au moins un arc, donc un sommet origine d'un arc), on peut alors construire un chemin de longueur arbitrairement grande. En effet, à chaque étape de la construction, on arrive

à un sommet S du graphe de valence entrante strictement positive (on arrive au sommet en question), et distinct des précédents (sinon le chemin construit contiendrait un circuit). Mais alors, comme par hypothèse $d_+(\mathrm{S}) = d_-(\mathrm{S})$, la valence sortante de S est elle aussi strictement positive, et donc on peut trouver un arc partant de S, c'est-à-dire que l'on peut continuer la construction.

Mais une telle construction n'est pas possible, attendu que le nombre de sommets du graphe orienté G est fini. En conclusion, on peut trouver dans G un circuit. La propriété 1.1 nous assure alors que G contient un circuit élémentaire, puisque tout circuit se décompose en circuits élémentaires.

Soit C un circuit élémentaire de G. Soit G' le graphe orienté obtenu en enlevant à G les arcs de C. Les sommets de G' ont pour valences sortantes et entrantes la même que pour G si C ne passe pas par eux, celle qu'ils ont pour G moins 1, si C passe par eux. Donc G' est encore pseudo-symétrique et possède au plus n arcs, et donc G' est réunion de circuits élémentaires $\mathrm{C}_1, \ldots, \mathrm{C}_p$ disjoints d'après l'hypothèse de récurrence.

De plus les arcs de C_1 (resp $\mathrm{C}_2, \ldots, \mathrm{C}_p$) sont des arcs de G', donc sont disjoints des arcs de C. Ainsi G est réunion des circuits élémentaires disjoints $\mathrm{C}, \mathrm{C}_1, \ldots, \mathrm{C}_p$.

On conclue alors par le principe de récurrence que tout graphe orienté pseudo-symétrique est réunion de circuits élémentaires disjoints.

□

1.3 Graphes Traversables

Définition Un graphe orienté G est fortement connexe si pour tous sommets S_1 et S_2 distincts, il existe un chemin du graphe orienté allant de S_1 à S_2.

Définition

- **(i)** Un circuit d'un graphe G est dit eulérien s'il comporte une et une seul fois chaque arc de G.
- **(ii)** Un graphe orienté G est dit traversable s'il existe un circuit eulérien de G. (*i.e.* un circuit passant une et une seule fois par chaque arc de G)

Théorème 1.3 *Un graphe orienté fortement connexe est traversable si et seulement si il est pseudo-symétrique.*

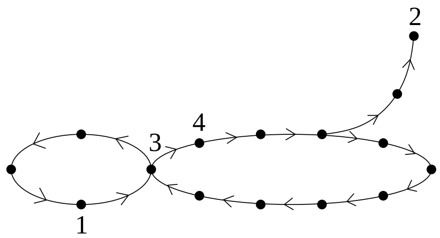
Le cas ou G ne possède qu'un seul sommet est trivial. G se réduit alors soit à un point, soit à un point et une boucle autour de ce point. Nous supposerons désormais que G possède au moins deux sommets.

Soit G un graphe orienté traversable, S un sommet de G. Les valences entrante et sortante de S sont égales au nombre d'occurences de S dans tout circuit eulérien de G ne partant pas de S (sinon, il faut enlever 1 à ces valences). Notons tout de même que comme G possède au moins deux sommets, il existe un circuit eulérien ne partant pas de S (un circuit eulérien passant par tous les sommets de G, il suffit de commencer ailleurs si le circuit de départ commence en S). La valeur des valences entrante et sortante du sommet S est bien commune, ce pour tout sommet, et donc le graphe orienté G est bien pseudo-symétrique.

Réciproquement, soit G un graphe orienté pseudo-symétrique et fortement connexe. D'après la propriété 1.2, on sait que G peut se décomposer en un ensemble de circuits élémentaires $C_1, \ldots, C_r$ disjoints.

Soit $r = 1$, auquel cas le graphe orienté G se réduit au circuit élémentaire C_1, et est bien traversable. Dans le cas contraire, C_1 a nécessairement un sommet en commun avec l'un des circuits C_i pour $i \neq 1$. En effet soit S_1 un sommet de C_1 et S_2 un sommet non atteint par C_1 (si un tel sommet n'existe pas, le problème est réglé). Comme G est supposé fortement connexe, il existe un chemin joignant S_1 à S_2. Soit S_3 le dernier sommet de ce chemin par lequel passe C_1, et S_4 le sommet suivant du chemin. Comme par construction C_1 ne passe pas par S_4, l'arc S_3S_4 n'appartient pas à C_1. Or les circuits $C_1, \ldots, C_r$ comprennent tous les arcs de G, et particulier l'arc S_3S_4, qui appartient par conséquent à un circuit C_i, avec $i \neq 1$. Mais alors C_1 et C_i ont bien un sommet en commun, et l'on peut les recoller.

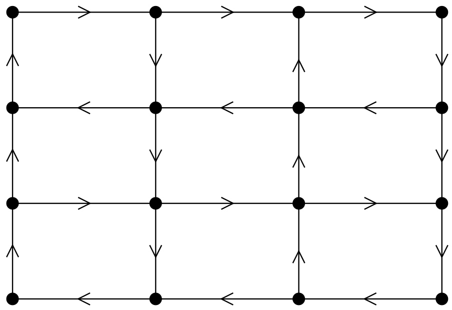
Exemple : Dans la figure qui suit, on considère le sommet 1, qui est sur le premier circuit élémentaire. Par connexité, on peut le joindre au sommet 2, sommet qui lui n'est pas sur le premier circuit (celui de gauche). Sur le chemin de l'un à l'autre, il faut bien que l'on sorte du premier circuit à un moment donné, ici entre le sommet noté 3 et le sommet noté 4. Mais alors l'arc $(3, 4)$ appartient à un autre circuit élémentaire que l'on recolle au premier de la manière évidente (on parcours un "huit").



Soit C'_1 le circuit obtenu en partant de S_3 et en parcourant successivement C_1 et C_i. On a obtenu une décomposition de l'ensemble des arcs de G en un ensemble de $r-1$ circuits disjoints, qui sont $C'1$ et les C_j pour j différent de 1 et i. Il suffit ensuite de recommencer le procédé, en tout $r-1$ fois, pour obtenir un circuit contenant tous les arcs de G, c'est-à-dire un circuit eulérien. Et donc G est bien traversable.

□

Exemple : Une application très classique de ce théorème est le problème dit du laitier découragé. Il doit desservir un quartier dont toutes les rues sont à sens unique, selon le plan suivant :



Il voudrait bien sûr déterminer un parcours passant une et une seule fois par chacune des rues. Malheureusement, on s'aperçoit aisément qu'en 8 carrefours, les nombres de rues qui arrivent et de rues qui partent sont différents. Le théorème 1.3 nous assure que le graphe n'est pas traversable. Il n'existe pas non plus de chemin (éventuellement ouvert, donc) passant par chaque rues : si un tel chemin existait, seuls les sommets de départ et d'arrivée pourraient avoir des valences entrante et sortante distinctes.

2 Application au problème du digicode

Éliminons d'emblée le cas $n = 1$. Il faut évidemment taper toutes les lettres de l'alphabet pour être sûr d'ouvrir la porte, soit un sésame de p lettres (et $p = p^1 + 1 - 1$, notre formule fonctionne donc toujours).

Soit G le graphe orienté dont les sommets sont les mots de $n - 1$ lettres sur notre alphabet à p lettres. Les arcs de G sont définis comme suit :

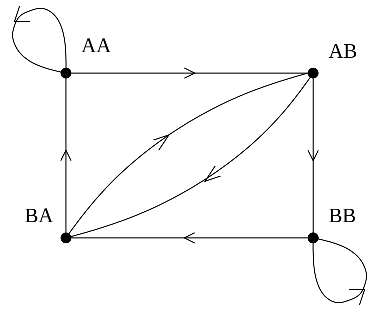
un arc relie le sommet S_1 au sommet S_2 si et seulement si il existe un mot de n lettres commençant par le mot S_1 et finissant par le mot S_2.

Par exemple, les mots 1234 et 2345 seront relié dans le cas d'un digicode "classique" où le mot recherché est composé de 5 caractères, sur l'alphabet $1, 2, 3, 4, 5, 6, 7, 8, 9, 0, A, B$. L'arc qui joint 1234 à 2345 correspond au mot 12345.

Exemple Donnons par exemple le graphe correspondant à un digicode à deux lettres, notées A et B, sur lequel on recherche un mot de trois lettres. Les mots possibles sont ici :

$$\mathrm{AAA}, \mathrm{AAB}, \mathrm{ABA}, \mathrm{ABB}, \mathrm{BAA}, \mathrm{BAB}, \mathrm{BBA}, \mathrm{BBB}$$

Ceci nous fait donc huit arcs pour le graphe cherché, dont les quatre sommets seront les mots de deux lettres, soit $\mathrm{AA}, \mathrm{AB}, \mathrm{BA}$ et BB.



On peut ici construire un circuit eulérien à la main, le graphe étant assez simple. Par exemple :

$$(\mathrm{BA}, \mathrm{AA}, \mathrm{AA}, \mathrm{AB}, \mathrm{BB}, \mathrm{BB}, \mathrm{BA}, \mathrm{AB}, \mathrm{BA})$$

Ce circuit correspond (un fois déplié) à la séquence de lettre BAAABBBABA, qui contient tous les mots de 3 lettres possibles. Elle comporte 10 lettres, soit $2^3 + 2$, comme prévu.

N.B. Le cas $n = 1$ pourrait aussi se traiter de cette façon, en considérant un graphe à un seul sommet (correspondant au mot vide), muni de p boucles correspondant aux p lettres...

Chaque sommet de G possède alors p arcs partant et p arcs arrivant : en effet il y a exactement p mots de n lettres commençant par un mot de $n-1$ lettre donné (il y a p lettres possibles pour la dernière), et de la même manière il y a p mots de n lettre se terminant par un mot de $n-1$ lettres donné. Dans le cas des mots dont toutes les lettres sont identiques, il y a une boucle, c'est-à-dire un arc qui joint le mot à lui-même. C'est le seul cas ou cela arrive, il n'affecte d'ailleurs en rien le décompte des arcs partants et arrivants en ce sommet : la boucle compte à la fois comme un arc partant et comme un arc arrivant.

En particulier les valences entrante et sortante de chaque sommet sont égales.

D'autre part le graphe orienté G est fortement connexe. Soient deux sommets S_1 et S_2 distincts donnés. Soient $u_1, \ldots, u_{n-1}$ la suite de lettres composant S_1 et $v_1, \ldots, v_{n-1}$ celle composant S_2. En les concaténant, on obtient un mot de $2n-2$ lettres, à savoir le mot $u_1, \ldots, u_{n-1}, v_1, \ldots, v_{n-1}$, qui définit un chemin allant de S_1 à S_2 dans le graphe G. (Les arcs sucessifs du chemin correspondent aux suites de n lettres consécutives de notre mot).

N.B. Ce chemin n'est pas nécessairement optimal. Par exemple, pour passer de ABABA à BABAB, on aura un chemin de longueur 5 alors qu'il y a ici un chemin de longueur 1. Pour trouver le plus court chemin, on procède de la façon suivante : soit q la taille de leur plus grande “intersection”, c'est-à-dire le plus grand indice tel que les suites $u_{n-q}, \ldots, u_{n-1}$ et $v_1, \ldots, v_q$ coïncident. Alors le mot $u_1, \ldots, u_{n-1}, v_{q+1}, \ldots, v_{n-1}$ définit le chemin le plus court qui joint S_1 à S_2. Il est de longueur $n-1-q$.

Le graphe orienté G vérifie donc les hypothèses du théorème 1.3, aussi nous pouvons affirmer qu'il est traversable, c'est-à-dire qu'il existe un circuit passant une et une seule fois par chacun de ses arcs.

Or à une tel circuit correspond un mot d'exactement $p^n + n - 1$ lettres, qui contient tous les mots de notre dictionnaire.

3 Résultats similaires et applications

Un résultat analogue au théorème 1.3 existe pour les graphes non orientés.

Pour distinguer ces nouveaux graphes des graphes orientés, on parlera cette fois d'arêtes du graphe, et non plus d'arcs. On définit la encore les notions de chemin, circuit et circuit élémentaire. On parle cette fois de valence tout court d'un sommet pour désigner le nombre de fois ou ce sommet est extrémité d'un arête (une boucle ayant ce sommet comme extrémités compte deux fois). Le résultat est alors le suivant :

Théorème 3.1 *Un graphe (non orienté) connexe est traversable si et seulement si tous ses sommets ont une valence paire.*

La démonstration de ce résultat est essentiellement la même que celle du théorème 1.3. Ici encore, l'une des implications est évidente : si un graphe traversable, le circuit trouvé nous montre que tous les sommets ont une valence paire.

Réciproquement, soit G un graphe connexe dont tous les sommets ont une valence paire. On commence par décomposer notre graphe en circuits disjoints : on construit un circuit “au hasard”, puis on enlève les arêtes par lesquels passe le circuit construit, et on recommence. Ensuite, grâce à la connexité du graphe, on peut concaténer les circuits obtenus en un seul.

On peut raffiner ce théorème pour obtenir le suivant :

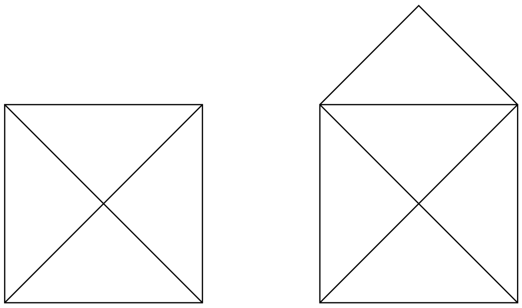
Théorème 3.2 *Soit* G *un graphe connexe. Il existe un chemin (et non plus un circuit) eulérien, c'est-à-dire passant une et une seule fois par chaque arête de* G*, si et seulement si le nombre de sommets ayant une valence impaire est égal à* 0 *ou* 2*.*

Bien sûr, s'il existe un chemin eulérien, le nombre de sommets ayant une valence impaire est 0 si le chemin est fermé, 2 si le chemin est ouvert.

Réciproquement, si le nombre de sommets ayant une valence impaire est 0, alors d'après le théorème précédent, G possède un circuit eulérien, donc un chemin eulérien. Si ce nombre est 2, nous allons construire un autre graphe.

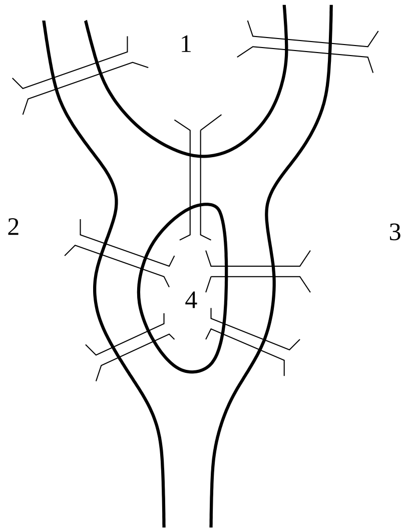
Soit G' le graphe obtenu en rajoutant à G un sommet, et deux arêtes, arêtes qui joignent respectivement le nouveau sommet aux deux sommet de valence impaire dans G. Alors G' est bien sûr toujours connexe (le nouveau sommet étant relié au reste du graphe). D'autre part, tous les sommets de G' ont cette fois une valence paire. Donc G' possède un circuit eulérien, circuit que l'on peut ouvrir à l'endroit du nouveau sommet pour obtenir un chemin eulérien de G.

Exemples :

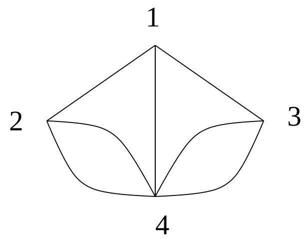


Le dessin de droite peut être effectué sans lever le crayon, pas celui de gauche. En effet dans celui de gauche, les quatre sommets ont une valence impaire, tandis que dans celui de droite, seuls les deux sommets du bas ont une valence impaire, et donc il existe un chemin eulérien.

D'autre part, l'adjectif “eulérien” provient bien sûr du mathématicien Euler, qui résolut en 1736 le problème suivant : est-il possible pour un piéton de la ville de Königsberg (désormais Kaliningrad) de traverser une et une seule fois chacun des 7 ponts de la ville ? La ville de Königsberg avait alors la structure suivante :



On peut représenter ce problème par le graphe suivant, où les ponts de Königsberg correspondent aux arêtes du graphe :



Là encore, les quatre sommets du graphe ont des valences impaires. Il n'est donc pas possible de le parcourir en passant une et une seule fois par chaque arête.

Références

[1] F. Harary, R.Z. Norman, D. Cartwright, *Introduction à la théorie des graphes orientés*, Dunod, 1968.

[2] C. Berge, *Graphes et hypergraphes*, Dunod, 1970.