



16

INTRODUCTION À LA THÉORIE DES GRAPHES

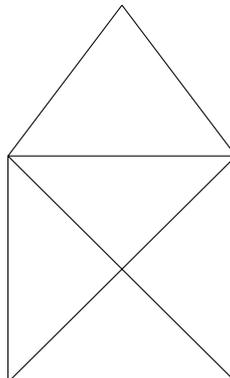
INTRODUCTION...

Introduits au milieu du XVIII^{ÈME} siècle par Euler, les graphes ont depuis trouvé de nombreuses applications dans des domaines variés comme la physique, la génétique; et leur intérêt s'est accru avec l'étude des réseaux de télécommunication, réseaux informatiques et des réseaux sociaux!

La théorie des graphes réunit sans mal les mathématiciens et informaticiens : autour du théorème des quatre couleurs par exemple... C'est actuellement un domaine très étudié en mathématiques discrètes.

Pour changer, une petite énigme... Peut-on, sur la figure ci-dessous :

- créer un chemin partant d'un sommet et en rejoignant un autre en passant une et une seule fois par chaque arête de la figure?
- créer un chemin fermé partant d'un sommet et revenant au même en passant une et une seule fois par chaque arête de la figure?



I DÉFINITIONS ET PREMIERS EXEMPLES

I.1 GRAPHE NON ORIENTÉ

DÉFINITIONS 1

SUR LES GRAPHS NON ORIENTÉS...

- D1** Un **graphe non orienté** \mathcal{G} est la donnée d'un couple (S, \mathcal{A}) , où :
- S est un ensemble fini, appelé ensemble des **sommets** de \mathcal{G} ,
 - \mathcal{A} est un ensemble de *paires* de sommets, appelé ensemble des **arêtes** de \mathcal{G} .
- D2** Les deux sommets définissant une arête sont ses **extrémités**.
- D3** Une **boucle** est une arête dont les deux extrémités sont identiques.
- D4** Un graphe est **simple** lorsqu'il ne contient ni boucle ni arête multiple entre deux mêmes sommets.
- D5** Un sommet est **isolé** lorsqu'aucune arête ne le relie à un autre sommet.
- D6** Deux sommets sont **adjacents** lorsqu'ils sont reliés par une arête.
- D7** L'**ordre** d'un graphe est son nombre de sommets.
- D8** Soit $s \in S$ un sommet de \mathcal{G} . Le **degré** de s , noté $d(s)$, est le nombre d'arêtes dont s est une extrémité.

En gros...

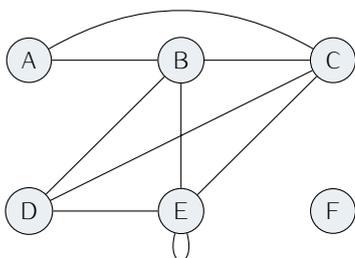
Une **paire** est un couple non ordonné (ou une partie à 2 éléments pouvant contenir deux fois le même élément) ; et un graphe est un ensemble de points reliés par des traits.

Petite remarque

Les boucles sont possibles car \mathcal{A} est un ensemble de paires, et pas un ensemble de parties à 2 éléments de S !

EXEMPLES 1

E1 On considère le graphe \mathcal{G} dont voici la représentation graphique :



- l'ensemble des sommets de \mathcal{G} est $\{A, B, C, D, E, F\}$, \mathcal{G} est d'ordre 6
- l'ensemble des arêtes de \mathcal{G} est :

$$\{A-B, A-C, B-C, B-D, B-E, C-D, C-E, D-E, E-E\}$$
- F est isolé
- $d(A) = 2, d(D) = 3, d(E) = 5$
- le graphe n'est pas simple : il possède une boucle sur E

E2 On pourrait modéliser le réseau social Facebook par un graphe non orienté sur lequel chaque compte est un sommet et la relation entre membres contacts se traduirait par une arête.

Attention !

Quand on cherche le degré, une boucle compte double !

Pour info...

En prenant comme données 4 milliards de comptes et, en moyenne, 338 amis par compte, ce graphe aurait $\frac{3 \times 10^9 \times 338}{2}$ arêtes...

PROPRIÉTÉS 1

FORMULE D'EULER

P1 Soient $n, p \in \mathbb{N}^*$ et \mathcal{G} un graphe à n sommets de degrés respectifs d_1, d_2, \dots, d_n et composé de p arêtes. On a :

$$\sum_{i=1}^n d_i = 2p$$

P2 Dans un graphe, il y a un nombre pair de sommets de degré impair.

Pour info...

Ce résultat est parfois appelé **lemme des poignées de mains**. Il affirme en effet que, dans une réunion de plusieurs personnes dont certaines se serrent la main, un nombre pair de personnes devra serrer un nombre impair de fois la main d'autres personnes.

* DÉMONSTRATION :

- P1.** Immédiat : chaque arête augmente de 2 la somme des degrés (soit elle augmente de 1 sur deux sommets distincts, soit de 2 sur un même sommet).
- P2.** Notons D_p la somme des degrés pairs et D_i la somme des degrés impairs. D'après P1 (et avec les notations utilisées) :

$$D_p + D_i = 2p$$

c'est à dire :

$$D_i = 2p - D_p$$

Or, D_p est une somme de nombres pairs, donc D_p est pair. Par conséquent : $2p - D_p$ est également pair, et donc D_i est pair.

Mais D_i est une somme de nombres impairs ; pour qu'elle soit paire, il est nécessaire qu'elle contienne un nombre pair de termes.

Conclusion : il y a un nombre pair de sommets de degré impair.

*

EXEMPLE 2

Les équipes de football des CPGE des lycées Ampère, ND des Minimes, Le Parc, Ste-Marie et Les Chartreux souhaitent organiser un tournoi dans lequel chaque équipe en affronterait trois autres. Que doit-on leur conseiller ?

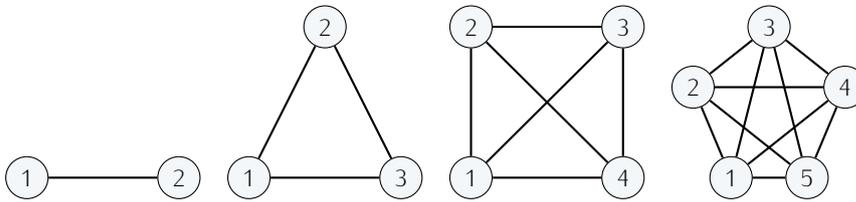
Pour terminer cette sous-partie sur les graphes non orientés :

DÉFINITION 2**GRAPHE COMPLET**

Un graphe **complet** est un graphe simple dont tous les sommets sont deux à deux adjacents.

EXEMPLE 3

Représentons les graphes complets d'ordres 2,3,4,5 :

**PROPRIÉTÉ 2**

Soit $\mathcal{G} = (S, \mathcal{A})$ un graphe d'ordre n .
Si \mathcal{G} est complet, alors il possède $\frac{n(n-1)}{2}$ arêtes.

*** DÉMONSTRATION :**

Voyons trois démonstrations de ce résultat...

1. Directement avec la définition.

Puisque \mathcal{G} est complet, en particulier :

- \mathcal{G} est simple, donc sans boucle ; et ainsi, toute arête est associée à un ensemble de sommets de cardinal 2 (deux sommets distincts) ;
- les sommets de \mathcal{G} sont deux à deux adjacents, et donc tout ensemble de sommets de cardinal 2 est associé à une arête.

Par conséquent, il y a autant d'arêtes dans \mathcal{G} que de sous-ensembles de S à deux éléments.

Or, $\text{Card}(S) = n$, donc il y a $\binom{n}{2}$ sous-ensembles de S à deux éléments.

Conclusion : \mathcal{G} possède $\frac{n(n-1)}{2}$ arêtes.

2. Avec la formule d'Euler.

Puisque \mathcal{G} est complet, chaque sommet de \mathcal{G} est adjacent à $n - 1$ sommets (tous les autres, sauf lui). Par conséquent, le degré de chaque sommet est $n - 1$.

On a alors, d'après la formule d'Euler (en notant p le nombre d'arêtes de \mathcal{G}) :

$$\sum_{i=1}^n (n-1) = 2p$$

Autrement dit :

$$n(n-1) = 2p$$

Conclusion : \mathcal{G} possède $\frac{n(n-1)}{2}$ arêtes.

3. Par récurrence...

- **Initialisation.** Pour $n = 1$:

Un graphe à un seul sommet ne possède pas d'arête... L'initialisation est ainsi vérifiée.

Important !

Résultat à connaître, mais dont la démonstration a été demandée dans le sujet **Ecritome 2024 Appli - Ex3...**

- **Hérédité.** Soit $n \in \mathbb{N}^*$. Supposons que "tout graphe d'ordre n possède $\frac{n(n-1)}{2}$ arêtes" et montrons que "tout graphe d'ordre $n+1$ possède $\frac{n(n+1)}{2}$ arêtes".

Soit $\mathcal{G} = (S, \mathcal{A})$ un graphe complet d'ordre $n+1$. Considérons $s_0 \in S$, notons $S' = S \setminus \{s_0\}$ ainsi que $\mathcal{G}' = (S', \mathcal{A}')$ le sous-graphe de \mathcal{G} induit par S' .

Par conséquent, \mathcal{G}' est un graphe complet d'ordre n et les arêtes de \mathcal{G} sont celles de \mathcal{G}' auxquelles il faut ajouter celles reliant s_0 aux autres sommets de \mathcal{G} . Autrement dit :

$$\mathcal{A} = \mathcal{A}' \cup \{s_0 - s \mid s \in S \text{ ET } s \neq s_0\}$$

Or, les deux ensembles de cette union étant disjoints (\mathcal{A}' ne contient, par définition, aucune arête dont s_0 est une extrémité), on a :

$$\begin{aligned} \text{Card}(\mathcal{A}) &= \text{Card}(\mathcal{A}') + \text{Card}(\{s_0 - s \mid s \in S \text{ ET } s \neq s_0\}) \\ &= \frac{n(n-1)}{2} + \text{Card}(\{s_0 - s \mid s \in S \text{ ET } s \neq s_0\}) \\ &= \frac{n(n-1)}{2} + n \\ &= \frac{n(n+1)}{2} \end{aligned}$$

) par hypothèse de récurrence
) n sommets de \mathcal{G} différents de s_0 , donc n arêtes de la forme $s - s_0$ car \mathcal{G} est complet

L'hérédité est ainsi vérifiée.

Conclusion : tout graphe complet d'ordre n possède $\frac{n(n-1)}{2}$ arêtes.

*

Autrement dit
 \mathcal{G}' est le graphe dont les sommets sont ceux de \mathcal{G} sauf s_0 , et dont les arêtes sont toutes celles de \mathcal{G} , sauf celles dont s_0 est une extrémité.

EXEMPLE 4

Lors d'une soirée, des étudiantes et étudiants de CPGE se saluent par un check. Tous les individus se sont salués, et on a compté 780 checks. Quel est l'effectif de la classe ?

I.2 GRAPHE ORIENTÉ

DÉFINITIONS 3
SUR LES GRAPHS ORIENTÉS...

D1 Un **graphe orienté** \mathcal{G} est la donnée d'un couple (S, \mathcal{A}) , où :

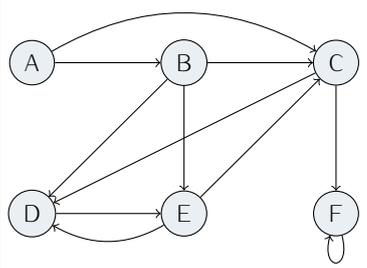
- S est un ensemble fini, appelé ensemble des **sommets** de \mathcal{G} ,
- \mathcal{A} est un ensemble de *couples* de sommets, appelé ensemble des **arcs** de \mathcal{G} .

D2 Soit $s \in S$ un sommet de \mathcal{G} . Le **degré** de s , noté $d(s)$, est le nombre d'arcs dont s est une extrémité. Le **degré entrant** de s , noté $d^-(s)$, est le nombre d'arcs entrant vers s ; et le **degré sortant** de s , noté $d^+(s)$, est le nombre d'arcs sortant de s .

Petite remarque
 On a alors : $d^-(s) + d^+(s) = d(s)$.

EXEMPLES 5

E1 On considère le graphe \mathcal{G} dont voici la représentation graphique :



- l'ensemble des sommets de \mathcal{G} est $\{A, B, C, D, E, F\}$, \mathcal{G} est d'ordre 6
- l'ensemble des arcs de \mathcal{G} est : $\{A \rightarrow B, A \rightarrow C, B \rightarrow C, B \rightarrow D, B \rightarrow E, C \rightarrow D, C \rightarrow F, D \rightarrow E, E \rightarrow C, E \rightarrow D, F \rightarrow F\}$
- $d^+(A) = 2, d^-(A) = 0, d^-(D) = 3, d^-(E) = 2, d^+(E) = 2$

E2 Le réseau social Instagram peut être modélisé par un graphe orienté.

E3 On peut modéliser le Web par un graphe orienté pour lequel :

- chaque sommet représente une page internet,

- chaque arc $i \rightarrow j$ traduit l'existence d'un lien sur la page i pointant vers la page j .

PROPRIÉTÉ 3

FORMULE D'EULER

Soient $n, p \in \mathbb{N}^*$ et \mathcal{G} un graphe à n sommets de degrés respectifs $d_1^-, d_1^+, d_2^-, d_2^+, \dots, d_n^-, d_n^+$ et composé de p arcs. On a :

$$\sum_{i=1}^n d_i^- = \sum_{i=1}^n d_i^+ = p$$

En particulier :

$$\sum_{i=1}^n d_i = 2p$$

Petite remarque

Le lemme des poignées de main reste valable en considérant le degré total des sommets, mais ne l'est plus si on distingue d^+ et d^- (voir Exemples 5 - E1).

* DÉMONSTRATION : Immédiat.

1.3 GRAPHE PONDÉRÉ

DÉFINITION 4

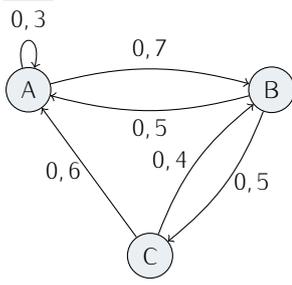
GRAPHE PONDÉRÉ

Un **graphe pondéré** est un graphe (orienté ou non) sur lequel chaque arc ou arête est pondéré par un réel strictement positif (appelé **poids**).

EXEMPLES 6

E1 Une carte routière peut se modéliser par un graphe pondéré dont les sommets sont les villes et les arêtes ou arcs sont les axes routiers indiquant les distances à parcourir sur chaque axe.

E2 On considère le graphe \mathcal{G} dont voici la représentation graphique :



Ce graphe peut modéliser l'évolution d'un objet pouvant se trouver en trois endroits différents A, B et C selon les règles suivantes :

- si l'objet est en A , il y reste avec une probabilité $0,3$, ou il va en B avec une probabilité $0,7$;
- si l'objet est en B , il va en A ou en C de façon équiprobable ;
- si l'objet est en C , il va en A avec une probabilité $0,6$ ou en B avec une probabilité $0,4$.

Petite remarque

On introduirait alors, pour tout $n \in \mathbb{N}$, a_n, b_n, c_n les probabilités que l'objet se situe respectivement en A, B, C ; et on aurait, pour tout $n \in \mathbb{N}$:

$$\begin{cases} a_{n+1} = 0,3a_n + 0,5b_n + 0,6c_n \\ b_{n+1} = 0,7a_n + 0,4c_n \\ c_{n+1} = 0,5b_n \end{cases}$$

système qu'on écrirait sous forme matricielle... Nous verrons cela avec l'étude des chaînes de Markov.

1.4 MATRICE D'ADJACENCE

DÉFINITION 5

MATRICE D'ADJACENCE D'UN GRAPHE

Soit $\mathcal{G} = (S, \mathcal{A})$ un graphe d'ordre n (non pondéré), et notons $S = \{s_1, s_2, \dots, s_n\}$.

On appelle **matrice d'adjacence** de \mathcal{G} la matrice $M = (m_{i,j}) \in \mathcal{M}_n(\mathbb{R})$ telle que pour tous $i, j \in \llbracket 1; n \rrbracket$:

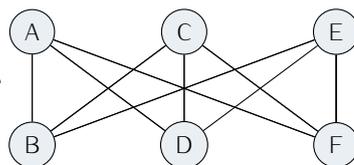
- $m_{i,j}$ est le nombre d'arêtes reliant s_i et s_j si \mathcal{G} est non orienté ;
- $m_{i,j}$ est le nombre d'arcs de s_i vers s_j si \mathcal{G} est orienté.

Important !

Par convention, les sommets seront rangés dans la matrice soit par ordre de numérotation soit par ordre alphabétique.

EXEMPLES 7

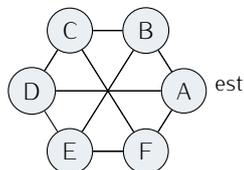
E1 La matrice d'adjacence du graphe



est la matrice

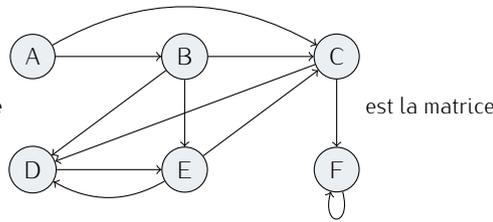
$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

E2 La matrice d'adjacence du graphe



Pour info...

Deux graphes peuvent avoir une allure différente mais la même matrice d'adjacence. On dit alors qu'ils sont **isomorphes**. Quand deux graphes sont isomorphes, leurs représentations graphiques sont identiques à déplacements des sommets près... Il y a alors correspondance unique entre un graphe et sa matrice d'adjacence, qui code alors toutes les informations du graphe. **C'est donc un outil qui doit nous permettre d'analyser le graphe en détail !** Affaire à suivre...



E3 La matrice d'adjacence du graphe

est la matrice

E4 La matrice d'adjacence d'un graphe non orienté est

E5 Un graphe est sans boucle si, et seulement si, sa matrice d'adjacence

E6 Un graphe non orienté est complet si, et seulement si, sa matrice d'adjacence

E7 Pour tout $i \in \llbracket 1; n \rrbracket$, la somme $\sum_{j=1}^n m_{i,j}$ est égale

E8 Pour tout $j \in \llbracket 1; n \rrbracket$, la somme $\sum_{i=1}^n m_{i,j}$ est égale

II DE L'ART DE RELIER DES SOMMETS...

II.1 CHÂÎNES ET CYCLES

DÉFINITIONS 6

CHÂÎNE, CYCLE, CYCLE EULÉRIEN

Soit $\mathcal{G} = (S, \mathcal{A})$ un graphe non orienté.

D1 Une **chaîne** est une liste finie et non vide de sommets telle que chaque paire de sommets consécutifs de la liste soit une arête de \mathcal{G} .

D2 La **longueur d'une chaîne** est le nombre d'arêtes dont elle est constituée.

D3 Une chaîne est **fermée** lorsque le sommet initial et le sommet final sont identiques.

D4 Une chaîne est **simple** lorsqu'aucune arête n'y figure plus d'une fois.

D5 Une **chaîne eulérienne** est une chaîne contenant *toutes les arêtes du graphe, chacune parcourue une seule fois*.

D6 Un **cycle** est une chaîne simple et fermée. Autrement dit, un cycle est une chaîne dans laquelle chaque arête n'est parcourue qu'une seule fois, et dont le départ et l'arrivée sont les mêmes.

D7 Un **cycle eulérien** est une chaîne eulérienne fermée.

Vocabulaire

Pour les graphes orientés, on parle de **chemin** plutôt que de chaîne et de **circuit** plutôt que de cycle. Les notions sont analogues, il suffit de faire attention au sens de parcours...

Attention !

Si une chaîne possède n sommets, sa longueur est $n - 1$.

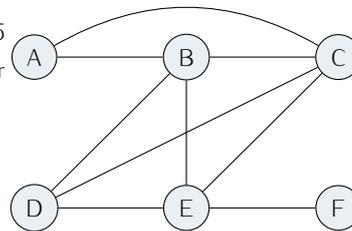
Petite remarque

En revanche, un sommet peut y figurer plusieurs fois.

EXEMPLE 8

Dans le graphe ci-contre :

- $A - B - C - D - E - F$ est une chaîne simple (pas fermée) de longueur 5
- $B - D - E - F - E - C - B$ est une chaîne fermée (pas simple) de longueur 6
- $A - B - C - D - E - F - E - C - B - A$ est une chaîne eulérienne
- Le graphe possède-t-il un cycle eulérien ?



Voyons l'information fournie par la matrice d'adjacence sur les chaînes ou chemins d'un graphe...

THÉORÈME 1

Soit $\mathcal{G} = (S, \mathcal{A})$ un graphe d'ordre n tel que $S = \{s_1, s_2, \dots, s_n\}$. On note M sa matrice d'adjacence.

Pour tout $p \in \mathbb{N}^*$ et tous $i, j \in \llbracket 1; n \rrbracket$, le coefficient (i, j) de la matrice M^p est égal

- soit au nombre de chaînes de longueur p reliant s_i et s_j si \mathcal{G} est non orienté,
- soit au nombre de chaînes de longueur p allant de s_i vers s_j si \mathcal{G} est orienté.

* DÉMONSTRATION : Démontrons ce résultat dans le cas d'un graphe non orienté, la démonstration est analogue dans le cas d'un graphe orienté.

Pour tout $p \in \mathbb{N}^*$ et tous $i, j \in \llbracket 1; n \rrbracket$, on note :

- $c_{i,j}(p)$ le nombre de chaînes de longueur p reliant les sommets i et j ,
- $m_{i,j}(p)$ le coefficient situé à la i -ème ligne et j -ième colonne de M^p .

Il s'agit donc d'établir :

$$\forall p \in \mathbb{N}^*, \forall i, j \in \llbracket 1; n \rrbracket, c_{i,j}(p) = m_{i,j}(p)$$

Démontrons ce résultat par récurrence !

- **Initialisation.** Pour $p = 1$:
Immédiat, par définition de la matrice d'adjacence d'un graphe et puisqu'une chaîne de longueur 1 est une arête.
- **Hérédité.** Soit $p \in \mathbb{N}^*$. Supposons : " $\forall i, j \in \llbracket 1; n \rrbracket, c_{i,j}(p) = m_{i,j}(p)$ " et montrons : " $\forall i, j \in \llbracket 1; n \rrbracket, c_{i,j}(p+1) = m_{i,j}(p+1)$ ".
Soient $i, j \in \llbracket 1; n \rrbracket$. On a :

$$\begin{aligned} c_{i,j}(p+1) &= \text{nombre de } (p+1)\text{-chaînes de la forme } s_i - \dots - s_j \\ &= \sum_{k=1}^n \text{nombre de } (p+1)\text{-chaînes de la forme } s_i - \dots - s_k - s_j \end{aligned}$$

Or, pour tout $k \in \llbracket 1; n \rrbracket$:

- * soit s_k et s_j ne sont pas adjacents, et il n'existe pas de $(p+1)$ -chaîne de la forme $s_i - \dots - s_k - s_j$;
- * soit s_k et s_j sont adjacents, et dans ce cas le nombre de $(p+1)$ -chaînes de la forme $s_i - \dots - s_k - s_j$ est ainsi égal au nombre de p -chaînes de la forme $s_i - \dots - s_k$ (c'est à dire $c_{i,k}(p)$) multiplié par le nombre d'arêtes reliant s_k et s_j (c'est à dire $m_{k,j}(1)$).

Dans les deux cas, le nombre de $(p+1)$ -chaînes de la forme $s_i - \dots - s_k - s_j$ est égal à $c_{i,k}(p) \times m_{k,j}(1)$.
Reprenons alors le calcul précédent :

$$\begin{aligned} c_{i,j}(p+1) &= \sum_{k=1}^n c_{i,k}(p) \times m_{k,j}(1) && \text{par hypothèse de récurrence} \\ &= \sum_{k=1}^n m_{i,k}(p) \times m_{k,j}(1) && \text{définition du produit matriciel} \\ &= m_{i,j}(p+1) \end{aligned}$$

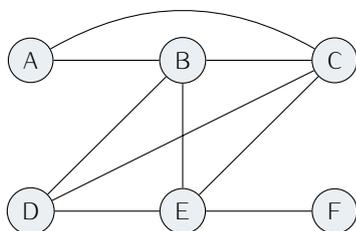
L'hérédité est ainsi vérifiée.

Conclusion : $\forall p \in \mathbb{N}^*, \forall i, j \in \llbracket 1; n \rrbracket, c_{i,j}(p) = m_{i,j}(p)$.

Petite remarque
Une chaîne de longueur 0 se réduit à un sommet... Par conséquent, le théorème est encore valable pour $p = 0$ avec la convention $M^0 = I_n$.

EXEMPLES 9

E1 Reprenons le graphe

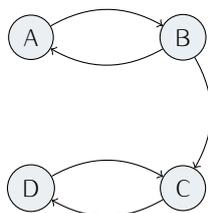


et notons M sa matrice d'adjacence.

On admet que $M^2 = \begin{pmatrix} 2 & 1 & 1 & 2 & 2 & 0 \\ 1 & 4 & 3 & 2 & 2 & 1 \\ 1 & 3 & 4 & 2 & 2 & 1 \\ 2 & 2 & 2 & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 4 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$ et $M^3 = \begin{pmatrix} 2 & 7 & 7 & 4 & 4 & 2 \\ 7 & 8 & 9 & 9 & 10 & 2 \\ 7 & 9 & 8 & 9 & 10 & 2 \\ 4 & 9 & 9 & 6 & 8 & 2 \\ 4 & 10 & 10 & 8 & 6 & 4 \\ 2 & 2 & 2 & 2 & 4 & 0 \end{pmatrix}$

- Le coefficient $(1, 2)$ de la matrice M^2 vaut 1 : il y a 1 chaîne de longueur 2 reliant A et B . Il s'agit de la chaîne $A - C - B$.
- Le coefficient $(2, 2)$ de la matrice M^2 vaut 4 : il y a 4 chaînes de longueur 2 reliant B à lui-même. Il s'agit des chaînes $B - A - B, B - C - B, B - D - B$ et $B - E - B$.
- Interprétons le coefficient $(4, 1)$ de la matrice M^3 :

E2 Considérons le graphe



et notons M sa matrice d'adjacence.

On admet que $M^2 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ et $M^9 = \begin{pmatrix} 0 & 1 & 0 & 4 \\ 1 & 0 & 5 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$

Rappels...
 ${}^t(MN) = ({}^tM)N$
Donc, pour tout $p \in \mathbb{N}$, ${}^t(M)^p = ({}^tM)^p$
D'où, si M est symétrique :
 ${}^t(M^p) = M^p$ et donc M^p est également symétrique.

- Le coefficient (1,1) de la matrice M^9 vaut 0 : il n'y a aucun chemin de longueur 9 reliant A à lui-même...
- Interprétons le coefficient (2,3) de la matrice M^9 :

II.2 CONNEXITÉ D'UN GRAPHE

DÉFINITIONS 7

GRAPHE CONNEXE

Soit \mathcal{G} un graphe.

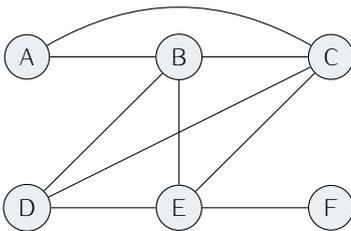
D1 Si \mathcal{G} est non orienté, on dit que \mathcal{G} est **connexe** lorsque chaque sommet de ce graphe peut être relié à n'importe quel autre sommet *par une chaîne*.

D2 Si \mathcal{G} est orienté,

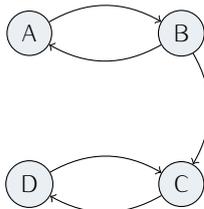
- on dit que \mathcal{G} est **faiblement connexe** lorsque son graphe non orienté associé est connexe,
- on dit que \mathcal{G} est **fortement connexe** lorsque chaque sommet de ce graphe peut être relié à n'importe quel autre sommet *par un chemin*.

Petite remarque
 Dans le cas d'un graphe orienté, si l'énoncé mentionne simplement "connexe", il s'agit en fait de la forte connexité.

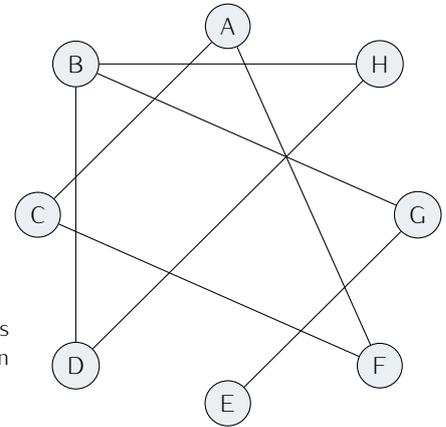
EXEMPLE 10



Ce graphe est connexe...



Ce graphe est faiblement connexe, mais pas fortement connexe (aucun chemin pour aller de C vers B).



Ce graphe n'est pas connexe : aucune chaîne reliant B et C...

Et voyons maintenant l'information fournie par la matrice d'adjacence sur la connexité d'un graphe...

THÉORÈME 2

Soit $\mathcal{G} = (S, \mathcal{A})$ un graphe d'ordre n tel que $S = \{s_1, s_2, \dots, s_n\}$. Notons M sa matrice d'adjacence.

1. Si \mathcal{G} est non orienté, alors : \mathcal{G} est connexe si, et seulement si, les coefficients de la matrice $I_n + M + \dots + M^{n-1}$ sont tous strictement positifs.
2. Si \mathcal{G} est orienté, alors : \mathcal{G} est fortement connexe si, et seulement si, les coefficients de la matrice $I_n + M + \dots + M^{n-1}$ sont tous strictement positifs.

* DÉMONSTRATION :

1. Supposons que \mathcal{G} est un graphe non orienté. Raisonnons par double implication.

\Leftarrow Supposons que les coefficients de la matrice $I_n + M + \dots + M^{n-1}$ sont tous strictement positifs, et montrons que \mathcal{G} est connexe.

Soient $i, j \in \llbracket 1; n \rrbracket$. Montrons que les sommets s_i et s_j sont reliés par une chaîne.

Par hypothèse, le coefficient (i, j) de la matrice $I_n + M + \dots + M^{n-1}$ est strictement positif. Par conséquent :

$$\exists p \in \llbracket 0; n-1 \rrbracket \mid (M^p)_{i,j} > 0$$

Et ainsi, d'après le théorème 1, il existe $p \in \llbracket 0; n-1 \rrbracket$ et au moins une chaîne de longueur p entre s_i et s_j .

Conclusion : il existe une chaîne entre s_i et s_j , d'où la connexité de \mathcal{G} .

\Rightarrow Supposons que \mathcal{G} est connexe et montrons que les coefficients de la matrice $I_n + M + \dots + M^{n-1}$ sont tous strictement positifs.

Soient $i, j \in \llbracket 1; n \rrbracket$. Montrons que le coefficient (i, j) de $I_n + M + \dots + M^{n-1}$ est strictement positif.

Rappel...
 Si une somme de termes est strictement positive, alors il existe au moins un terme strictement positif (quelle est la contraposée de cette assertion?).

- ★ On sait déjà que les coefficients des matrices $I_n, M, M^2, \dots, M^{n-1}$ sont positifs ou nuls. Par conséquent, pour montrer que le coefficient (i, j) de la matrice $I_n + M + \dots + M^{n-1}$ est strictement positif, il suffit de montrer qu'il existe $p \in \llbracket 0; n-1 \rrbracket$ tel que le coefficient (i, j) de M^p est strictement positif.
- ★ Or, \mathcal{G} est connexe. Ainsi, il existe une chaîne reliant s_i et s_j . Autrement dit, il existe $p \in \mathbb{N}$ et une chaîne de longueur p reliant s_i et s_j .
D'après le théorème 1, on en déduit :

$$\exists p \in \mathbb{N} / (M^p)_{i,j} > 0$$

Il ne reste plus qu'à justifier que l'on peut en fait trouver p dans $\llbracket 0; n-1 \rrbracket$... C'est bien le cas, puisque pour aller de s_i à s_j par une chaîne, il est inutile de passer deux fois par le même sommet ! La chaîne la plus courte (elle existe bien...) pour aller de s_i à s_j est donc de longueur inférieure ou égale à $n-1$.

Conclusion : le coefficient (i, j) de $I_n + M + \dots + M^{n-1}$ est strictement positif.

2. Analogue au cas précédent.

★

Pourquoi ?

Car les coefficients de M sont positifs...

II.3 EXISTENCE DE CHAÎNE EULÉRIENNE ET DE CYCLE EULÉRIEN

THÉORÈME 3

D'EULER

Soit \mathcal{G} un graphe simple non orienté et connexe.

1. \mathcal{G} possède un cycle eulérien si, et seulement si, tous ses sommets sont de degré pair.
2. \mathcal{G} possède une chaîne eulérienne si, et seulement si, son nombre de sommets de degré impair est 0 ou 2. Dans le cas où \mathcal{G} possède deux sommets de degré impair, alors ces deux sommets sont les extrémités de la chaîne eulérienne.

Vocabulaire

On dit parfois qu'un graphe est eulérien lorsqu'il possède au moins un cycle eulérien.

★ DÉMONSTRATION : A chercher, pour celles et ceux qui veulent s'occuper un peu.

★

III DEUX ALGORITHMES CLASSIQUES

III.1 ALGORITHME DE PARCOURS EN PROFONDEUR (HP)

L'objectif de l'algorithme de parcours en profondeur d'un graphe \mathcal{G} est de déterminer les *composantes connexes* d'un graphe.

Commençons par l'exploration en profondeur d'un graphe à partir d'un sommet initialement choisi.

Soient alors $\mathcal{G} = (S, \mathcal{A})$ et s un sommet de \mathcal{G} .

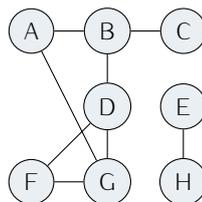
- on commence par visiter s ,
- puis, pour chaque sommet adjacent à s qui n'a pas encore été visité, noté t :
 - ★ on visite t ,
 - ★ puis, pour chaque sommet adjacent à t qui n'a pas encore été visité, noté u :
 - ◇ on visite u ,
 - ◇ puis ...

On obtient ainsi une liste de tous les sommets qu'il est possible de relier à s .

A ce stade, il est possible que des sommets n'aient pas été visités : c'est le cas si le graphe n'est pas connexe. On peut relancer l'algorithme en initialisant sur un sommet non visité.

EXEMPLE 11

Détaillons l'algorithme de parcours en profondeur sur le graphe



Partons initialement du sommet A ($L=[A]$), puis on visite tous les sommets non visités adjacents à A :

- on visite donc B ($L=[A,B]$), puis on visite tous les sommets non visités adjacents à B :
 - ★ on visite donc C ($L=[A,B,C]$), puis on visite tous les sommets non visités adjacents à C : il n'y en a pas (on ne poursuit donc pas en profondeur ici, on remonte...)
 - ★ on visite donc D ($L=[A,B,C,D]$), puis on visite tous les sommets non visités adjacents à D :
 - ◇ on visite donc F ($L=[A,B,C,D,F]$), puis on visite tous les sommets non visités adjacents à F :
 - ▷ on visite donc G ($L=[A,B,C,D,F,G]$), puis on visite tous les sommets non visités adjacents à G : tous ont déjà été visités !

Petite remarque

Nous voici donc en face d'un algorithme récursif, que nous aurons le plaisir de programmer en **Python**.

Attention !

Cette liste n'a aucune raison d'être une chaîne...

Petite remarque

J'écris, au fur et à mesure, la liste L des sommets déjà visités.

A la fin, on obtient donc la liste $[A, B, C, D, F, G]$ qui est la liste de tous les sommets qui peuvent être reliés à A par une chaîne.

Vocabulaire
On parle de **composante connexe contenant A** du graphe (vocabulaire HP).

III.2 ALGORITHME DE DIJKSTRA

L'objectif de l'algorithme de Dijkstra est de déterminer *une chaîne de poids minimal* entre deux sommets quelconques d'un graphe pondéré (orienté ou non).

Considérons $\mathcal{G} = (S, \mathcal{A})$ un graphe (orienté ou non) et notons, pour tous $u, v \in S$ ($u \neq v$) $p_{u,v}$ le poids de l'arête $u - v$ (ou de l'arc $u \rightarrow v$) si cette arête existe; on considère que $p_{u,v} = +\infty$ si l'arête $u - v$ n'existe pas.

Petite remarque
Si le graphe n'est pas pondéré, il permet alors de déterminer le plus court chemin (et sa longueur) entre deux sommets...

Fixons $s \in S$ un sommet et déterminons, pour tout $t \in S$, la distance de s à t , notée $d(s, t)$, c'est à dire la longueur du plus court chemin reliant s à t .

A chaque étape de l'algorithme, nous allons noter $\ell(s, t)$ la longueur du plus court chemin alors exhibé reliant s à t ; si un tel chemin n'a pas encore été exhibé, on posera $\ell(s, t) = +\infty$.

- Initialement, on considère alors :

$$\ell(s, s) = 0 ; \forall t \in S \setminus \{s\}, \ell(s, t) = +\infty$$

- A chaque étape, on sélectionne le sommet u qui contient la plus petite valeur de $\ell(s, u)$ et pour chacun de ses sommets adjacents v (n'ayant pas déjà été sélectionné), on effectue le test : "est-ce qu'il est plus court de passer par u pour aller à v ou non?" Autrement dit, cela revient à tester la condition :

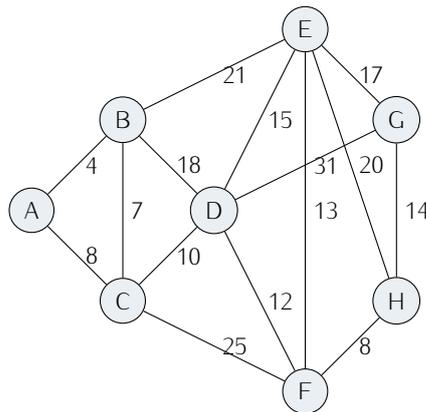
$$\ell(s, u) + p_{u,v} \stackrel{?}{\leq} \ell(s, v)$$

- * Si non, on ne modifie rien et on poursuit.
- * Si oui, on remplace l'ancienne valeur de $\ell(s, v)$ par $\ell(s, u) + p_{u,v}$ en précisant bien le sommet u de provenance.

Petite remarque
A chaque étape de l'algorithme, on choisit le sommet dont la distance à s est minimale... L'algorithme de Dijkstra est donc un algorithme glouton.

EXEMPLE 12

Détaillons l'algorithme de Dijkstra pour déterminer les plus courts chemins (et leur longueur) reliant le sommet A aux autres sommets du graphe ci-dessous :



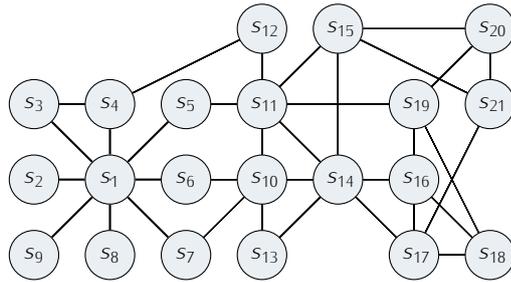
A	B	C	D	E	F	G	H	sommet choisi
0	∞	∞	∞	∞	∞	∞	∞	A
	<u>4_A</u>	8 _A	∞	∞	∞	∞	∞	B
		<u>8_A</u>	22 _B	25 _B	∞	∞	∞	C
			<u>18_C</u>	25 _B	33 _C	∞	∞	D
				<u>25_B</u>	30 _D	49 _D	∞	E
					<u>30_D</u>	42 _E	45 _E	F
						42 _E	<u>38_F</u>	H
						<u>42_E</u>		G

Important !
A chaque étape, on regarde si la somme de la distance menant au sommet sélectionné + celle vers ses sommets adjacents est inférieure à la distance déjà écrite pour aller vers ce-dit sommet adjacent. Si oui, on indique cette nouvelle valeur en précisant de quel sommet elle est obtenue... Puis on souligne la valeur minimale de chaque ligne pour indiquer le nouveau sommet sélectionné. On raye ensuite toutes les colonnes du dessous pour indiquer que le plus court chemin du sommet initial au sommet sélectionné a été trouvée.

- Pour trouver le plus court chemin reliant A à G, il suffit alors de prendre la colonne G : on récupère l'information 42_E indiquant que la longueur est de 42 et qu'il faut venir de E; puis dans la colonne de E, on voit qu'il faut venir de B; et dans la colonne de B, on voit qu'il faut venir de A...
Le plus court chemin pour aller de A à G est le chemin $A - B - E - G$.
- Le plus court chemin pour aller de A à H est le chemin $A - C - D - F - H$.

IV ÉLÉMENTS D'ANALYSE DE RÉSEAUX SOCIAUX (HP)

La dernière partie est l'occasion de se focaliser davantage sur l'analyse des réseaux sociaux, et plus particulièrement sur la recherche d'*influenceurs*. Pour cela, considérons le graphe suivant :



PREMIER INDICATEUR : CENTRALITÉ DE DEGRÉ

La première approche, un peu naïve, serait de dire que l'influenceur le plus important est le sommet de plus haut degré. On l'appelle **centralité de degré**, définie par :

$$C_D(i) = d(i)$$

En clair : "plus j'ai d'amis, plus je suis influent"

C'est un premier indicateur, mais il est nettement insuffisant. En effet, sur l'exemple précédent, le sommet s_1 a le degré le plus élevé, mais les sommets qui lui sont adjacents ont tous un degré égal à 1 ou faible...

En gros...
 T'as plein d'amis, mais tes amis ont peu d'amis : tu n'es pas un influenceur.

SECOND INDICATEUR : CENTRALITÉ DE PROXIMITÉ (OU CENTRALITÉ HARMONIQUE)

Commençons pour cela par donner un indicateur de proximité entre deux sommets : la **distance** entre deux sommets distincts s_i et s_j est la longueur de la plus courte chaîne reliant s_i à s_j . Si aucune chaîne ne les relie, on dit que la distance entre les deux est infinie.

Notons $\Delta_{i,j}$ la distance entre les sommets s_i et s_j . Avec pour convention $\frac{1}{\infty} = 0$, on définit maintenant le degré de proximité du sommet s_i par le nombre :

$$C_P(i) = \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{\Delta_{i,j}}$$

Plus les distances entre s_i et les autres sommets sont petites, plus $C_P(i)$ sera grand.

En clair : "plus je suis proche d'un grand nombre de personnes, plus je suis influent"

Pour info...
 La quantité $\frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}} \Delta_{i,j}$ pourrait aussi convenir si le graphe était connexe... S'il ne l'est pas, on trouverait 0 pour tous les sommets.

TROISIÈME INDICATEUR : CENTRALITÉ D'INTERMÉDIARITÉ

Avec ce troisième indicateur, on souhaite mesurer l'importance d'un sommet dans le passage rapide d'une information. Autrement dit, si un sommet est fréquent dans un gros nombre de chaînes les plus courtes entre deux autres sommets, on veut que son degré d'intermédiation soit fort.

Notons alors $v_{j,k}$ le nombre de plus courtes chaînes reliant s_j et s_k , ainsi que $v_{j,k}(i)$ le nombre de plus courtes chaînes reliant s_j et s_k passant par i .

On définit maintenant le degré d'intermédiation du sommet s_i par le nombre :

$$C_I(i) = \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{\substack{k=1 \\ k \neq i, k \neq j}}^n \frac{v_{k,j}(i)}{v_{k,j}}$$

En clair : "plus vous passez par moi, plus je suis influent"

QUATRIÈME INDICATEUR : CENTRALITÉ SPECTRALE

A travers ce dernier indicateur, on souhaite mesurer l'influence d'un sommet en tenant compte de l'influence de ses sommets adjacents...

En clair : "plus j'ai d'amis influents, plus je le suis"

Ce qui nous amène à définir la centralité spectrale de façon implicite... Supposons le graphe simple et notons $M = (m_{i,j})_{1 \leq i,j \leq n}$ sa matrice d'adjacence. On a alors :

$$\forall (i, j) \in \llbracket 1; n \rrbracket^2, m_{i,j} = \begin{cases} 1 & \text{si } s_i \text{ et } s_j \text{ sont adjacents} \\ 0 & \text{sinon} \end{cases}$$

Notons $C_S(i)$ la centralité spectrale du sommet s_i , nombre que l'on souhaite positif et que l'on définit par l'existence d'un réel strictement positif λ tel que :

$$C_S(i) = \frac{1}{\lambda} \sum_{j \text{ tq } s_j \text{ adj à } s_i} C_S(j)$$

$$= \frac{1}{\lambda} \sum_{j=1}^n m_{i,j} C_S(j)$$

On voit bien le caractère implicite dans cette définition : on a besoin des indicateurs sur les voisins de s_i pour calculer l'indicateur sur s_i ... Mais pour calculer l'indicateur d'un voisin de s_i , on aura besoin de l'indicateur de ses voisins, dont s_i fait partie ! On tourne en rond... sauf si l'on remarque que l'on a alors :

$$\sum_{j=1}^n m_{i,j} C_S(j) = \lambda C_S(i)$$

et qu'en notant $X = \begin{pmatrix} C_S(1) \\ C_S(2) \\ \vdots \\ C_S(n) \end{pmatrix}$, on obtient :

$$MX = \lambda X$$

Autrement dit : le vecteur colonne des centralités spectrales des sommets du graphe est un vecteur propre de M pour une certaine valeur propre de M .

Bon, il reste tout de même deux petits soucis :

- Comment choisit-on λ parmi les valeurs propres de M ?
- Concrètement, comment fait-on pour calculer ce vecteur ?

Pour répondre à ces deux questions, on combine deux résultats importants :

1. le théorème de Perron-Frobenius,
2. la méthode de la puissance itérée...

Amazing !!

N'est-ce pas plein de beauté ça ?!
On comprend mieux le nom de cette centralité...

Un peu d'histoire

- Oskar Perron (1880-1975, allemand) doit presque toute sa postérité à ce fameux théorème, même si sa contribution ne se limite pas qu'à celui-ci...
- Ferdinand Georg Frobenius (1849-1917, allemand) a établi d'importants résultats en théorie des groupes et en algèbre linéaire.

ES Pour info...

L'algorithme **PageRank** de Google, qui trie les pages du Web selon leur pertinence/importance, est basé sur cette centralité spectrale...
Voir ESSEC 2008 E 2.