

Théorie des graphes

Université de Namur
Renaud Lambiotte, Lionel Tabourier

2013-2014

Table des matières

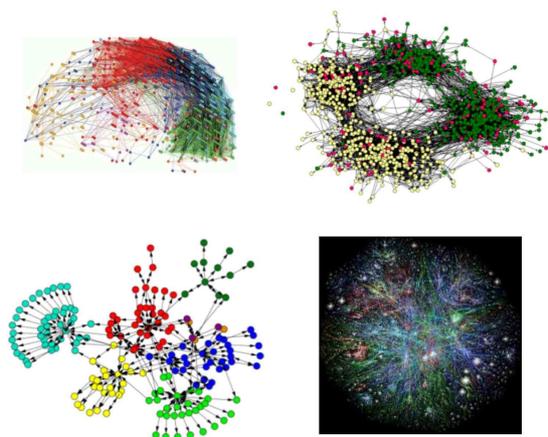
1	Introduction	3
2	Graphes connexes et eulériens	4
2.1	Euler et ses ponts	4
2.2	Graphes, sommets, arêtes, degré	4
2.3	Parcours, connexité et graphes eulériens	6
2.4	Le théorème d'Euler	6
3	Plus courts chemins et centralité	9
3.1	Notations matricielles	9
3.2	Parcours et chemins	10
3.3	Graphes bipartis	11
3.4	Mesures de centralité	12
3.5	Algorithme de Dijkstra	15
4	Arbres et connectivité	18
4.1	Arbres, forêts et sous-arbres	18
4.2	Caractérisation des arbres	20
4.3	Nombre d'arbres sous-tendants	21
4.4	Problème de l'arbre sous-tendant de poids minimum	24
4.5	Connectivité	26
4.6	Théorème de Menger	27
4.7	Un réseau de communication fiable	28
4.8	Notion de robustesse	29
5	Graphes hamiltoniens	31
5.1	Chemins et cycles hamiltoniens	31
5.2	Condition nécessaire	32
5.3	Conditions suffisantes	32
5.4	Graphes eulériens, graphes hamiltoniens et problèmes d'algorithmique	33
6	Mariages, couplages et couvertures	35
6.1	Définitions de base	35
6.2	Couplages maximaux, chemins alternés et chemins augmentés	35
6.3	Couplage de graphes bipartis	37
6.4	Problème d'assignation des tâches	40
6.5	Parenthèse sur les systèmes de recommandation	41

7	Coloriages d'arêtes	42
7.1	Définitions de base	42
7.2	Indice chromatique et degré maximum	43
7.3	Relation avec les couplages	46
7.4	Problème de la confection d'horaire	46
8	Cliques, ensembles indépendants et l'impossible désordre	47
8.1	Le théorème de l'amitié	47
8.2	Ensembles indépendants et couvertures de sommets et cliques	47
8.3	Théorème de Ramsey	49
8.4	Cliques et densité	52
8.5	Théorie de Ramsey	52
9	Coloriage de sommets	53
9.1	Définitions de base	53
9.2	Polynôme chromatique	54
10	Graphes planaires	56
10.1	Question du coloriage d'une carte	56
10.2	Notion de graphe planaire	57
10.3	Graphes non-planaires	57
10.4	Retour au coloriage d'une carte	58

Chapitre 1

Introduction

Dans le monde qui nous entoure, de nombreux systèmes peuvent être représentés par des réseaux, où des éléments sont en interaction les uns avec les autres. On peut penser au cerveau humain, composé d'une myriade de neurones connectés par des synapses, au Web où des pages sont connectées par des hyperliens, l'Internet composés de routeurs ou les réseaux sociaux tels Facebook. La possibilité d'analyser des systèmes de nature différente par le même ensemble d'outils réserve une place à la théorie des graphes et des réseaux dans la boîte à outils des mathématiciens.



Le but de ce cours est de donner une introduction aux algorithmes de base de la théorie des graphes, pour répondre à des questions telles que : comment trouver son chemin dans des grands graphes ou comment y trouver des sommets importants. Ce cours se veut rigoureux, avec une approche mathématique du problème et des démonstrations des résultats, pour vous permettre de comprendre la logique derrière les outils présentés et vous donner les bases pour en découvrir d'autres à l'avenir.

Evaluation :

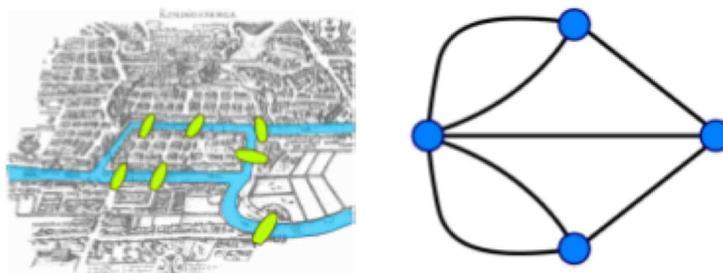
L'évaluation se fera par un examen écrit en janvier. La moitié des points sera attribuée à des exercices du même genre que ceux traités au cours et proposés en séances de travaux dirigés. L'autre moitié des points à la restitution de définitions, énoncés de théorèmes et démonstrations vues au cours.

Chapitre 2

Graphes connexes et eulériens

2.1 Euler et ses ponts

Leonhard Euler, né à Bâle en 1707, est considéré comme l'un des plus grands mathématiciens. Parmi ses nombreuses contributions, il est à l'origine d'un des premiers travaux en théorie des graphes. En 1735, lors d'une visite à Königsberg (aujourd'hui Kaliningrad), il entend parler d'une question qui taraude les bourgeois de la ville : *Peut-on se promener dans la ville en empruntant chacun des sept ponts une et une seule fois, et revenir à son point de départ ?*



Euler étudie le problème et démontre un théorème qui permet de répondre à cette question pour n'importe quelle ville. Pour ce faire, il représente la ville de manière abstraite, avec des sommets connectés par des arêtes, c'est-à-dire par des graphes. Comme nous verrons plus bas, la question devient alors : Existe-t-il un parcours qui passe par chaque arête une et une seule fois ? Le graphe est-il eulérien ?

2.2 Graphes, sommets, arêtes, degré

Avant de nous lancer dans la résolution du problème des ponts de Königsberg, nous allons introduire certaines notions et concepts de base de théorie des graphes.

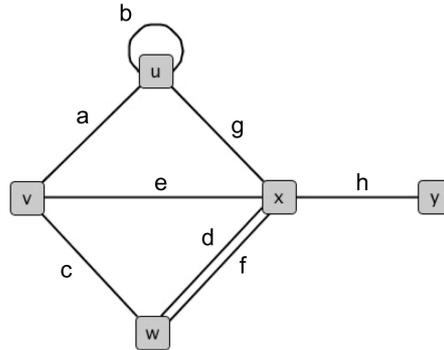
* *Définition*

Un **graphe** est un triplet (V, E, ψ) tel que :

- V est un ensemble dont les éléments sont appelés **sommets** ou **nœuds**,
- E est un ensemble dont les éléments sont appelés **arêtes**,

- ψ est une fonction, dite **fonction d'incidence**, qui associe à chaque arête un sommet ou une paire de sommets.

Toute la structure du graphe est représentée par ces trois ingrédients : l'ensemble des sommets, l'ensemble des arêtes et les relations entre ces deux ensembles.



Par exemple, dans l'exemple ci-dessous, le graphe G est décrit par le triplet (V_G, E_G, ψ_G) . L'ensemble des sommets est

$$V_G = \{u, v, \dots, y\}, \quad (2.1)$$

l'ensemble des arêtes est

$$E_G = \{a, b, \dots, g\}, \quad (2.2)$$

et la fonction d'incidence, définie sur l'ensemble des arêtes, donnée par

$$\begin{aligned} \psi_G(a) &= uv \\ \psi_G(b) &= uu \\ &\dots \\ \psi_G(g) &= ux. \end{aligned} \quad (2.3)$$

* *Définition*

Définitions générales sur les graphes :

- Deux sommets incidents à la même arête sont dits **adjacents**.
- Une arête incidente à un seul sommet est une **boucle**.
- Le nombre de sommets d'un graphe G est noté $\gamma(G)$, et celui d'arêtes $\epsilon(G)$.
- Le **degré** d'un sommet est le nombre d'arêtes incidentes à celui-ci, une boucle compte pour deux.
- Un **sous-graphe** du graphe (V, E, ψ) est un graphe (V', E', ψ') tel que $V' \subseteq V$, $E' \subseteq E$ et ψ' est la restriction de ψ à E' .

Dans l'exemple précédent, l'arête b est ainsi une boucle. Le degré sommet x vaut 5, celui de u vaut 4, etc.

Il est important de noter que la représentation du graphe sous forme de diagramme, toute utile soit-elle pour mieux comprendre la structure du système, est parfois trompeuse. Un même graphe peut être représenté de toutes sortes de manières différentes et pourtant toujours garder la même structure. Dans ce cas, on dira que les graphes sont isomorphes l'un à l'autre. Même si leurs sommets et leurs arêtes ont des noms différents, et leur représentation graphique diffère, leur structure est identique. De manière formelle, on dit que deux graphes G et G' , caractérisés par (V', E', ψ') sont isomorphes si il existe des bijections

$$f : V \rightarrow V'$$

$$g : E \rightarrow E' \quad (2.4)$$

telles que

$$\psi(e) = uv \text{ ssi } \psi'(g(e)) = g(u)g(v). \quad (2.5)$$

Vérifier que deux graphes sont isomorphes est une tâche complexe. Néanmoins, pour montrer que deux graphes ne le sont pas, il suffit parfois de vérifier que certaines propriétés diffèrent entre deux graphes : leur nombre de sommets différent, leur nombre d'arêtes, le nombre de sommets de degré 1, etc...

* *Définition*

Pour terminer, deux dernières définitions :

- Un **graphe planaire** est un graphe qui peut être représenté sous la forme de diagramme où les arêtes ne s'intersectent pas.
- Un **graphe simple** est un graphe qui ne comporte pas de boucles ni d'arêtes multiples entre les sommets.

2.3 Parcours, connexité et graphes eulériens

* *Définition*

Définitions générales sur les parcours :

- Un **parcours** est une suite $v_0e_1v_1e_2\dots e_nv_n$ où $v_0v_1\dots$ sont des sommets et $e_1e_2\dots$ sont des arêtes reliant les sommets consécutifs du parcours.
- La **longueur** d'un parcours est son nombre d'arêtes. Le sommet d'**origine** est v_0 , le sommet de **destination** est v_n . Les autres sommets sont dits **intérieurs**. Un parcours est **fermé** si $v_0 = v_n$.
- Un **chemin** est un parcours dont les sommets sont tous distincts.
- Un **cycle** est un parcours fermé dont les sommets d'origine et intérieurs sont tous distincts.
- Un graphe est **connexe** si pour chaque paire de points il existe un parcours qui les relie. Les **composantes connexes** d'un graphe sont des sous-graphes connexes maximaux.
- Un **parcours est eulérien** s'il visite chaque arête une et une seule fois. Un **graphe est eulérien** s'il existe un parcours eulérien fermé.
- Une **piste** est un parcours qui passe au maximum une fois par ses arêtes.
- Un parcours eulérien est donc une piste fermée qui passe par toutes les arêtes du graphe.

2.4 Le théorème d'Euler

• **Théorème**

Un graphe connexe est eulérien si et seulement si tous ses sommets sont de degré pair.

Par le théorème d'Euler, il n'existe donc pas de parcours eulérien fermé dans le graphe de Königsberg. Qu'en est-il de Namur, par contre ?

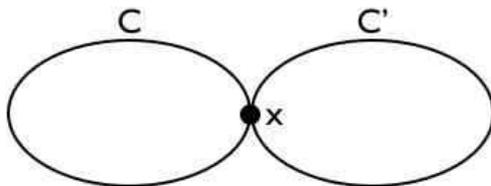
il est toujours possible de quitter b si on y arrive. $d^{\text{eff}}(b)$ diminue de deux unités et reste pair. On continue jusqu'à ce que l'on arrive sur a , auquel cas une piste fermée a été construite. Etant donné que le graphe est fini, ce retour se fait en un nombre fini d'étapes.

Revenons-en à la démonstration du théorème d'Euler.

◦ **Preuve**

Soit G un graphe eulérien et C un parcours eulérien de G , partant du sommet u . A chaque fois que v est visité, $d^{\text{eff}}(v)$ diminue de deux unités. Comme un parcours eulérien comporte toutes les arêtes de G , $d(v)$ est pair pour tout $v \neq u$. De la même manière, comme C commence et finit en u , $d(u)$ est aussi pair. $\rightarrow G$ n'a pas de sommet de degré impair.

Soit G un graphe dont tous les sommets sont de degré pair. Supposons que G ne soit pas eulérien. Par le lemme précédent, on sait qu'il existe des pistes fermées dans G . Considérons la plus grande piste fermée de G . Comme G n'est pas eulérien, $G - E(C)$ comporte des arêtes. Soit G' une composante connexe dans $G - E(C)$. Dans G' , tous les degrés sont pairs. Choisissons un sommet x de G' qui appartienne aussi à C . Un tel sommet existe parce que G est connexe. Par le lemme, il existe une piste fermée C' partant de x dans G' .



Le parcours CC' est fermé et ne parcourt aucun lien plus d'une fois. Or $|E(CC')| > |E(C)|$, ce qui contredit le fait que C est la plus grande piste fermée de G et montre par l'absurde que G est eulérien.

• **Théorème**

Existence d'un parcours eulérien : Un graphe connexe possède un parcours eulérien si et seulement si le nombre de nœuds de degré impair est 0 ou 2.

La preuve est similaire au théorème précédent. Si le nombre de sommets de degré impair est égal à 2, il existe un parcours eulérien joignant ces deux sommets.

Chapitre 3

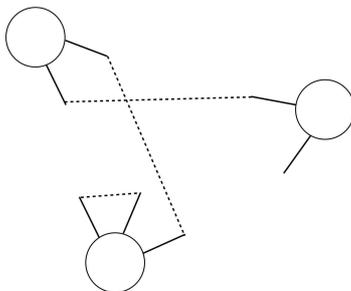
Plus courts chemins et centralité

3.1 Notations matricielles

- **Théorème**

Théorème des poignées de mains : La somme des degrés des nœuds d'un graphe est deux fois le nombre d'arêtes.

Un corrolaire de ce théorème est qu'il n'est pas possible de trouver, dans un graphe, un seul sommet avec un degré impair. Pour s'en convaincre, considérons un graphe dont le degré des sommets serait 3, 2 et 2. Une arête d'un sommet peut être vue comme une demi-arête, qui doit être complétée par une autre demi-arête. Pour cette raison, la somme des degrés doit être un nombre pair, sinon il reste une demi-arête qui n'est pas complétée.



Pour démontrer ce résultat, il est pratique d'adopter une approche matricielle pour représenter le graphe.

- * **Définition**

La **matrice d'adjacence** de G est la matrice carrée $\gamma \times \gamma$ dont l'élément ij est le nombre d'arêtes entre le sommet v_i et le sommet v_j .

- * **Définition**

La **matrice d'incidence** de G est la matrice rectangulaire $\gamma \times \epsilon$ dont l'élément ij est le nombre de fois (i.e., 0, 1 ou 2 pour une boucle) que le sommet v_i est incident à l'arête e_j .

- **Preuve**

Soit M la matrice d'incidence de G . Cette matrice vérifie $\sum_i M_{ij} = 2$ et $\sum_j M_{ij} = d(i)$. On a donc que

$$\sum_{ij} M_{ij} = \sum_i \left(\sum_j M_{ij} \right) = \sum_i d(i) \quad (3.1)$$

et

$$\sum_{ij} M_{ij} = \sum_j \left(\sum_i M_{ij} \right) = 2\epsilon(G), \quad (3.2)$$

ce qui montre le théorème.

• **Théorème**

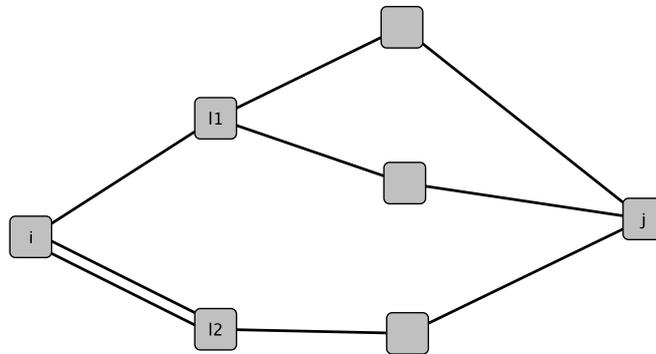
Nombre de parcours d'un graphe : Soit A la matrice d'adjacence d'un graphe, alors l'élément ij de A^k (avec $k \geq 0$) est le nombre de parcours de longueur k de v_i vers v_j .

◦ **Preuve**

Démonstration par récurrence. Vrai pour $k = 1$: A_{ij} est le nombre d'arêtes entre i et j et donc le nombre de parcours de longueur 1 entre ces sommets. On suppose que le théorème est vrai pour k , et on démontre qu'il est vrai pour $k + 1$.

$$(A^{k+1})_{ij} = (AA^k)_{ij} = \sum_l A_{il}(A^k)_{lj}. \quad (3.3)$$

$A_{il}(A^k)_{lj}$ est le nombre de parcours de longueur $k + 1$ passant par l . En sommant sur tous les sommets intermédiaires possibles, on compte donc tous les chemins de longueur $k + 1$ entre i et j .



Dans l'exemple précédent, par exemple, on a

$$\begin{aligned} (A^2)_{l_1j} &= 2, \\ (A^2)_{l_2j} &= 1, \\ (A^2)_{il_1} &= 1, \\ (A^2)_{il_2} &= 2, \end{aligned} \quad (3.4)$$

et on trouve bien que le nombre total de parcours de longueur 3 entre i et j vaut 4.

3.2 Parcours et chemins

* *Définition*

La distance $d(u, v)$ entre les sommets u et v d'un graphe est le nombre d'arêtes minimal d'un parcours entre ces 2 sommets. Si un tel parcours n'existe pas, $d(u, v) = \infty$.

• **Lemme**

Si $P = u \dots u' \dots v' \dots v$ est un parcours de longueur minimale entre u et v , alors le sous-parcours $Q = u' \dots v'$ est un parcours de longueur minimale entre u' et v' .

◦ **Preuve**

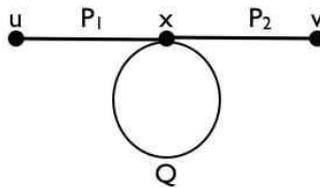
Par l'absurde. Par définition, on décompose P en P_1QP_2 . Supposons qu'il existe un parcours Q' entre u' et v' et strictement plus court que Q . Alors $P' = P_1Q'P_2$ est strictement plus court que P , ce qui est contradiction avec l'hypothèse, et démontre le lemme.

• **Lemme**

Un plus court parcours est un chemin \Rightarrow On parlera alors de plus court chemin.

◦ **Preuve**

Par l'absurde. On suppose que ce n'est pas vrai et que le plus court parcours passe deux fois par un sommet x .



Le parcours Q de x à x n'est pas le plus court. Contradiction par le lemme précédent.

3.3 Graphes bipartis

* **Définition**

Un graphe est biparti si il existe une partition des sommets en deux ensembles V_1 et V_2 telle que les sommets de V_1 ne sont adjacents qu'aux sommets de V_2 et vice-versa.

• **Théorème**

Un graphe est biparti si et seulement si il ne contient pas de cycle de longueur impaire.

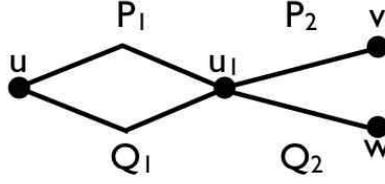
◦ **Preuve**

G est biparti avec bipartition (X, Y) et soit $C = v_0v_1 \dots v_kv_0$ un cycle de longueur $k + 1$. On suppose que $v_0 \in X$. Puisque $v_0v_1 \in E$ et G est biparti, $v_1 \in Y$. En general, on a $v_{2i} \in X$ et $v_{2i+1} \in Y \forall i \geq 0$. Comme $v_0 \in X$, on a $v_k \in Y$, et donc $k = 2i + 1$. La longueur du cycle est paire.

Soit G un graphe connexe sans cycle de longueur impaire. On prend un sommet arbitraire u . On définit la partition (X, Y) de la manière suivante :

$$\begin{aligned} X &= \{x \in V \mid d(u, x) \text{ est paire}\} \\ Y &= \{y \in V \mid d(u, y) \text{ est impaire}\} \end{aligned} \tag{3.5}$$

On va montrer qu'il n'y a aucune arête entre deux sommets de X . Soit v et w deux sommets de X . On considère P un plus court chemin de u à v , et Q un plus court chemin de u à w . Soit u_1 le dernier sommet en commun de P et Q .



P_1 et Q_1 ont la même longueur. Les longueurs de P et Q sont des nombres pairs. Les longueurs de P_2 et Q_2 ont donc la même parité. Considérons le parcours $P_2^{-1}Q_2$. Sa longueur est un nombre pair. De plus, il s'agit d'un chemin entre v et w car aucun de ses sommets n'est visité deux fois. Si il existe une arête entre v et w , il existe un cycle de longueur impaire. Contradiction. Les sommets de X ne sont pas adjacents les uns aux autres.

3.4 Mesures de centralité

Lors de l'analyse de grands graphes, une question primordiale est d'identifier des sommets jouant un rôle plus important que d'autres. Dans ce but, de nombreuses mesures de centralité ont été introduites. Soit un graphe défini par sa matrice d'adjacence A_{ij} symétrique. La mesure de centralité la plus simple est tout simplement le degré d'un sommet.

$$k_i = \sum_j A_{ij} \quad (3.6)$$

L'importance d'un sommet est proportionnelle à son nombre d'arêtes.

Dans cette définition, chaque arête contribue avec le même poids à la centralité d'un sommet. Dans de nombreuses situations, cependant, ce n'est pas le cas. Par exemple, lorsque vous cherchez un travail, une lettre de recommandation de Bill Gates vaut certainement plus que la mienne. C'est précisément pour cette raison que la notion de *eigenvector centrality* a été définie : une personne est importante si elle est connectée à de nombreuses personnes importantes. Mathématiquement, cela s'écrit

$$x_i = \frac{1}{\lambda} \sum_j A_{ij} x_j \quad (3.7)$$

où λ est une constante. Par définition, x_i est un vecteur propre de A_{ij} , et λ sa valeur propre associée. La centralité d'un sommet est donc proportionnelle à la somme des centralités de ses voisins. Par souci de simplicité, assumons que le graphe est connexe. En imposant la contrainte que les centralités sont des nombres positifs, on peut montrer par le théorème de Frobenius Perron que le vecteur de centralités est le vecteur propre dominant de la matrice d'adjacence, c'est-à-dire celui dont la valeur propre est la plus grande. En pratique, comment peut-on calculer ce vecteur propre dominant ? Le moyen le plus simple est d'utiliser la méthode de la puissance. Pour faire simple, supposons que λ est connu. A partir de la condition initiale $x_j(0) = 1/\gamma$, on itère

$$x_i(t) = \frac{1}{\lambda} \sum_j A_{ij} x_j(t-1), \quad (3.8)$$

ce qui nous donne

$$x_i(t) = \frac{1}{\lambda^t} \sum_j A_{ij}^t x_j(0). \quad (3.9)$$

Etant donné que la matrice A_{ij} est symétrique, ses vecteurs propres forment une base orthonormée. Classons les vecteurs propres $v^{(\alpha)}$ de la valeur propre la plus grande λ_1 à la plus petite λ_n , on peut réécrire

$$A_{ij} = \sum_{\alpha} \lambda_{\alpha} v_i^{(\alpha)} v_j^{(\alpha)}, \quad (3.10)$$

et il en découle

$$A_{ij}^t = \sum_{\alpha} \lambda_{\alpha}^t v_i^{(\alpha)} v_j^{(\alpha)}. \quad (3.11)$$

La récursion

$$x_i(t) = \sum_{\alpha} \left(\frac{\lambda_{\alpha}}{\lambda_1}\right)^t v_i^{(\alpha)} \sum_j v_j^{(\alpha)} x_j(0). \quad (3.12)$$

est donc assurée de converger vers le vecteur propre dominant de la matrice d'adjacence étant donné que $(\frac{\lambda_{\alpha}}{\lambda_1})^t \rightarrow 0$ quand $t \rightarrow \infty$ pour $\alpha > 1$. A noter que les autres vecteurs propres sont des modes qui se relaxent d'autant plus lentement que leur valeur propre est proche de la valeur propre dominante.

L'équation (3.9) montre clairement que l'eigenvector centralité dépend de toute la structure du graphe, et des parcours qui y sont présents (puissance de la matrice d'adjacence). Il existe d'autres types de mesures de centralité qui se focalisent seulement sur un type de parcours, les chemins les plus courts. Une mesure standard est la closeness centrality qui est définie comme la distance moyenne entre un sommet et tous les autres sommets du graphe

$$x_i^C = \frac{1}{N-1} \sum_{j \neq i} d(i, j). \quad (3.13)$$

Un sommet est central si il tend à être proche des autres sommets. Une autre mesure basée sur la notion de chemin le plus court est la betweenness centrality, définie pour un sommet i comme la fraction des chemins les plus courts entre n'importe quelle paire de sommets qui passe par i

$$x_i^B = \sum_{j \neq k \neq i} \frac{\sigma(j, k, i)}{\sigma(j, k)}, \quad (3.14)$$

où $\sigma(j, k)$ est le nombre de chemins les plus courts entre j et k et $\sigma(j, k, i)$ celui de tels chemins passant par i . La betweenness centrality est une mesure de l'importance d'un sommet pour le flot d'information, dans l'hypothèse que l'information suit les chemins les plus courts.

Ce petit intermède sur la notion de centralité nous amène naturellement à deux méthodes qui se sont révélées extrêmement influentes pour mesurer l'importance d'une page Web. Celles-ci font écho à la notion de eigenvector centrality, avec comme différence principale qu'elles sont dédiées aux graphes dirigés.

* *Définition*

Un graphe dirigé, ou digraphe, est un triplet (V, E, ψ) où

- V est l'ensemble des sommets ;
- E est l'ensemble des arêtes ;
- ψ est la fonction d'incidence, qui associe à chaque arête une paire ordonnée de sommets.

Dans un graphe dirigé, les arêtes ont une direction (comme des hyperliens) et la matrice d'adjacence est en générale asymétrique. Par définition A_{ij} indique la présence ou l'absence d'arête allant de j vers i . Pour de tels graphes, on peut définir le degré entrant, comme le nombre total d'arêtes dirigées vers un sommet, $k_i^{in} = \sum_j A_{ij}$ et le degré sortant, comme le nombre d'arêtes quittant un sommet, $k_i^{out} = \sum_j A_{ji}$. Le Pagerank d'un sommet est défini par la relation

$$x_i = \sum_j \frac{A_{ij}}{k_j^{out}} x_j, \quad (3.15)$$

qui est simplement la solution stationnaire d'une chaîne de Markov sur le graphe dirigé. En effet, l'équation

$$x_i(t) = \sum_j \frac{A_{ij}}{k_j^{out}} x_j(t-1), \quad (3.16)$$

donne l'évolution de trouver un marcheur aléatoire sur un sommet du graphe. A chaque étape, un marcheur choisit une arête autour de lui au hasard. Par construction, la matrice de transition vérifie la condition

$$\sum_i \frac{A_{ij}}{k_j^{out}} = 1, \quad (3.17)$$

associée à la conservation de la probabilité à chaque étape. La principale différence avec l'eigenvector centrality est que la somme est pondérée : des sommets ayant de nombreuses arêtes sortantes contribuent moins à l'importance d'un sommet que ceux en ayant peu. Pour en revenir à l'exemple de la lettre de recommandation, une lettre de Bill Gates a plus de valeur si elle est unique, plutôt que si il en écrit une tous les jours. Pagerank peut se calculer en utilisant la méthode de la puissance sur de très grandes matrices. En octobre 2002, Google battait ainsi le record de l'étude de la matrice la plus grande, avec plus de 3 milliards de sommets. Il est intéressant de noter que Pagerank se réduit au degré lorsque le graphe est non-dirigé. Dans la pratique, le graphe du Web n'est pas connexe, comme nous l'avons implicitement assumé, et la convergence de (3.16) n'est pas assurée. De surcroit, l'état stationnaire n'est pas en général unique. Pour cette raison, des trucs ont été utilisés pour rendre la dynamique ergodique : les marcheurs peuvent se téléporter à travers le graphe.

Une autre mesure d'importance de sommets dans des graphes dirigés différencie deux types de rôle. L'algorithme HITS de Kleinberg définit pour chaque sommet un score de hub x et un score de authority y . De manière circulaire, il définit un bon hub comme un sommet qui pointe vers de bonnes autorités, et une bonne autorité comme un sommet qui reçoit de nombreuses connections de bons hubs. Mathématiquement, les scores de hub et de authority sont définis comme les solutions asymptotiques du système

$$\begin{aligned} x_i(t) &= \sum_j A_{ji} y_j(t-1) \\ y_i(t) &= \sum_j A_{ij} x_j(t), \end{aligned} \quad (3.18)$$

ou, de manière équivalente, du système

$$\begin{aligned} x_i(t) &= \sum_j D_{ij} x_j(t-1) \\ y_i(t) &= \sum_j D_{ji} y_j(t-1). \end{aligned} \quad (3.19)$$

La matrice $D_{ij} = \sum_k A_{ki} A_{jk}$ est souvent appelée la matrice d'autorité. Sa transposée est appelée matrice de hub. Une fois de plus, la notion de centralité est reliée à la valeur propre dominante d'une matrice associée au graphe. Nous ne reviendrons pas sur les propriétés spectrales des graphes dans ce cours, mais il faut savoir qu'elles trouvent de nombreuses applications, par exemple pour le clustering de graphes.

3.5 Algorithme de Dijkstra

Dans cette section, nous allons introduire un algorithme efficace pour trouver rapidement des chemins le plus court dans un graphe. Une solution efficace à ce problème est relativement récente. Elle est due à Edsger Wybe Dijkstra (1930-2002), célèbre, outre pour ses contributions en informatique, pour ses aphorismes : "L'informatique n'est pas plus la science des ordinateurs que l'astronomie n'est celle des télescopes."

* *Définition*

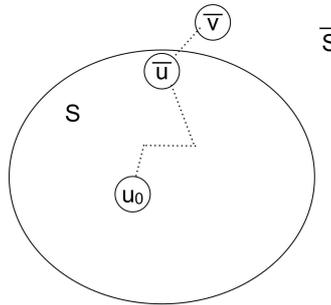
Les concepts du chapitre précédent peuvent être généralisés au cas de graphes pondérés.

- Une fonction de poids sur un graphe (V, E, ψ) est une fonction w de $E \rightarrow R$, qui associe un nombre réel à chaque arête. Un tel graphe est un graphe pondéré.
- Le poids, ou longueur d'un parcours est la somme des poids des arêtes qui le composent.
- La distance $d(u, v)$ est la longueur du plus court chemin de u à v .

Dans la suite, on va supposer que tous les poids sont strictement positifs. De nouveau, on parlera de plus court chemin car tout parcours le plus court est un chemin.

L'algorithme de Dijkstra permet de trouver le plus court chemin d'un sommet u_0 à un sommet v dans un graphe pondéré. Ce faisant, il trouve les plus courts chemins à tous les autres sommets du graphe.

Soit S un sous-ensemble de V tel que $u_0 \in S$ et $\bar{S} = V \setminus S$. Soit $P = u_0 \dots \bar{u}\bar{v}$ le plus court chemin de u_0 à \bar{S} , c'est-à-dire le plus court chemin entre u_0 et l'élément de \bar{S} qui en est le plus proche.



Alors il faut que $\bar{u} \in S$ et on a que $u_0 \dots \bar{u}$ est un plus court chemin entre u_0 et \bar{u} . De plus tout ce chemin est inclus dans S . Cela implique que

$$\begin{aligned} d(u_0, \bar{v}) &= d(u_0, \bar{u}) + w(\bar{u}\bar{v}) \\ d(u_0, \bar{v}) &= \min_{u \in S, v \in \bar{S}} \{d(u_0, u) + w(uv)\} \end{aligned} \quad (3.20)$$

Cette formule est à la base de l'algorithme de Dijkstra. On va partir de $S_0 = \{u_0\}$ et construire des séquences $S_0, S_1, S_2, \dots, S_i, \dots$ de sous-ensembles de V de sorte que, à chaque étape i , les plus courts chemins entre u_0 et tous les sommets de S_i sont connus.

L'idée de l'algorithme de Dijkstra est la suivante.

1. Déterminer le sommet le plus proche de u_0 . On calcule $d(u_0, \bar{S}_0)$ en cherchant le sommet u_1 tel que $d(u_0, u_1) = \min_{x \in \bar{S}_0} d(u_0, x) = \min_{x \in \bar{S}_0} w(u_0x)$
2. $S_1 = \{u_0, u_1\}$. Soit $P_1 = u_0u_1$ le plus court chemin entre u_0 et u_1 . On cherche $u_2 \in \bar{S}_1$ tel que $d(u_0, u_2) = \min_{x \in \bar{S}_1} d(u_0, x)$. Par (3.20), on sait que ce chemin le plus court est composé d'un plus court chemin dans S_1 , auquel on ajoute une arête.

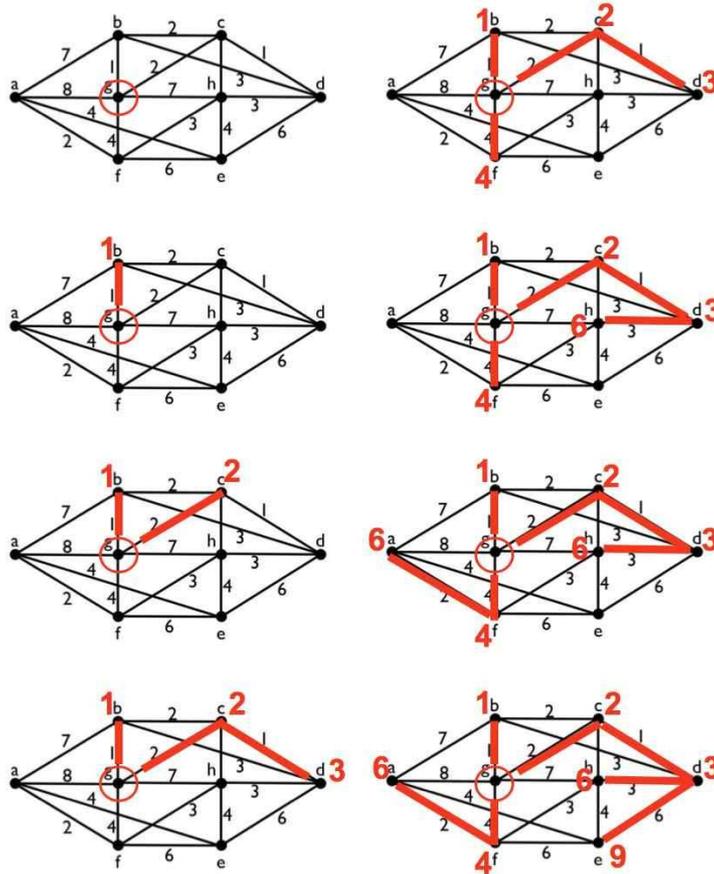
3. En général, soit l'ensemble $S_k = \{u_0, u_1, \dots, u_k\}$ et les plus courts chemins P_1, P_2, \dots, P_k . Par construction, chacun de ces plus courts chemins est entièrement dans S_k . On calcule $d(u_0, \overline{S}_k)$ en utilisant

$$d(u_0, \overline{S}_k) = \min_{u \in S_k, v \in \overline{S}_k} \{d(u_0, u) + w(uv)\} \quad (3.21)$$

et on définit $u_{k+1} \in \overline{S}_k$ comme le sommet qui minimise cette distance. Ce faisant, on trouve la distance entre u_0 et u_{k+1}

$$d(u_0, u_{k+1}) = d(u_0, u_j) + w(u_j u_{k+1}), \quad (3.22)$$

pour un certain $j \leq k$, et le plus court chemin en ajoutant l'arête $u_j u_{k+1}$ au chemin P_j .



A chaque étape, on indique la distance à u_0 des sommets dans S_k , et on indique les chemins les plus courts en rouge. A noter que, au final, l'ensemble des arêtes rouges forme un sous-graphe sans cycle. Ce que l'on appelle un arbre.

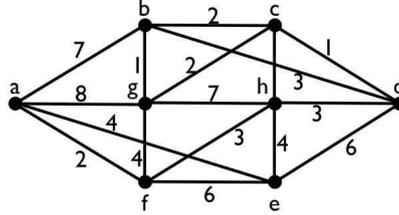
L'algorithme de Dijkstra est un raffinement de cet algorithme, qui permet d'éviter de faire des opérations inutiles. En effet, à chaque étape, on recalcule des distances à des éléments pour lesquels cette information est connue. Pour éviter ces répétitions, on associe à chaque sommet un nombre $l(v)$ qui est une borne supérieure de $d(u_0, v)$, et on procède comme suit.

1. $l(u_0) = 0$ et $l(v) = \infty \forall v \neq u_0$, $S_0 = \{u_0\}$ et $i = 0$.
2. Pour chaque $v \in \overline{S}_i$, remplacer $l(v)$ par le minimum de $\{l(v), l(u_i) + w(u_i v)\}$. Il est important de noter que pour calculer $l(u_i) + w(u_i v)$, on ne doit considérer que les voisins de u_i dans \overline{S}_i . Chercher le v tel que $l(v)$ est minimum, et l'appeler u_{i+1} . $S_{i+1} = S_i \cup \{u_{i+1}\}$,

3. Si $i = \gamma - 1$, on arrête. Sinon, $i \rightarrow i + 1$ et retourner au point 2.

L'algorithme de Dijkstra s'implémente souvent sous la forme de tableau, comme dans l'exemple qui suit.

exemple: On cherche les plus courts chemins de g aux autres sommets du graphe.



S	a	b	c	d	e	f	g	h
g	∞	∞	∞	∞	∞	∞	0	∞
g,b	8	1	2	∞	∞	4		7
g,b,c	8		2	6	∞	4		7
g,b,c,d	8			3	∞	4		7
g,b,c,d,f	8				9	4		6
g,b,c,d,f,h	6				9			6
g,b,c,d,f,h,a	6				9			
g,b,c,d,f,h,a,e					9			

A noter que l'algorithme peut être interrompu plus tôt si l'on est intéressé que par le chemin le plus court vers un sommet particulier. L'algorithme de Dijkstra est quadratique, en ce sens qu'il se termine en un temps de l'ordre γ^2 : γ itérations et au maximum $\gamma - 1$ distances à évaluer à chaque itération.

L'algorithme de Dijkstra se généralise facilement au cas de graphes dirigés.

Chapitre 4

Arbres et connectivité

La SNCB va mal, l'entretien des rails pèse gravement sur son budget. Le conseil d'administration décide de supprimer des lignes, autant que possible, tout en préservant la connectivité du réseau. Que tout un chacun puisse malgré tout se déplacer à travers la Belgique. Dans cette optique, quelle la structure optimale du réseau belge ? Celle d'un arbre !



4.1 Arbres, forêts et sous-arbres

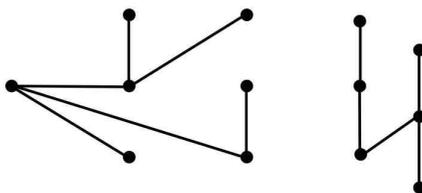
* *Définition*

Un **arbre** est un graphe connexe et sans cycle.

* *Définition*

Une **forêt** est un graphe sans cycle.

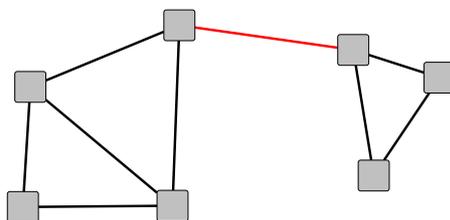
exemple:



* *Définition*

Une arête de G est fragmentante si le nombre de composantes connexes augmente lorsque cette arête est retirée $W(G - e) > W(G)$.

exemple:

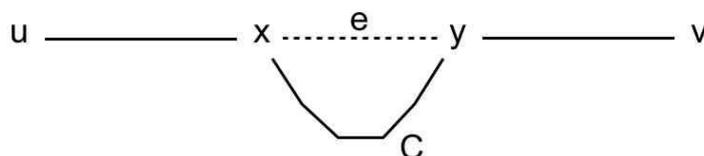


• Lemme

Une arête e de G est fragmentante ssi elle n'appartient à aucun cycle de G .

◦ **Preuve**

→ Soit e , entre les sommets x et y , une arête fragmentante de G . Comme $W(G - e) > W(G)$, il existe une paire de sommets u et v telle qu'il existe un chemin de u à v dans G , mais pas dans $G - e$. Ce chemin passe nécessairement par e .



Il existe un chemin entre u et x et un chemin entre y et v dans $G - e$. Si e appartenait à un cycle C , x et y seraient connectés par $C - e$ dans $G - e$. Il existerait un chemin entre u et v dans $G - e$, ce qui contredit l'énoncé du théorème.

← Soit une arête e , entre les sommets x et y , qui n'appartient à aucun cycle et n'est pas fragmentante. Il en découle que $W(G - e) = W(G)$. Etant donné que x et y sont reliés par e dans G , ils sont dans la même composante connexe de G . x et y sont donc dans la même composante connexe de $G - e$. Il existe un chemin P entre x et y dans $G - e$. Il en découle que l'arête e appartient au cycle $P + e$ dans G . Contradiction.

• Lemme

Un graphe connexe est un arbre ssi toute arête est fragmentante

◦ **Preuve**

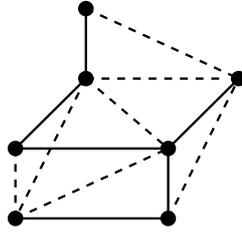
→ Soit G un arbre et e une arête. Comme G est acyclique, e n'appartient à aucune arête. Par le lemme précédent, e est une arête fragmentante de G .

← Soit un graphe G connexe dont toute arête est fragmentante. Supposons que G n'est pas un arbre. G contient un cycle C . Par le lemme précédent, aucune arête de C n'est fragmentante. Contradiction.

* **Définition**

Un **sous-graphe sous-tendant** ou **couvrant** d'un graphe G est un sous-graphe qui contient tous les sommets de G .

exemple:



• **Théorème**

Tout graphe connexe contient un arbre sous-tendant.

◦ **Preuve**

Soit G connexe et T un sous-graphe sous-tendant minimal de G , c'est-à-dire que si on lui retire une arête, T n'est plus connexe. $W(T) = 1$ et $W(T - e) > 1$ pour toute arête de T . Toute arête de T est fragmentante. Par le lemme précédent, T est un arbre.

4.2 Caractérisation des arbres

• **Théorème**

Soit G un graphe à n sommets et m arêtes, alors les propositions suivantes sont équivalentes :

1. G est connexe et sans cycle (G est un arbre)
2. G est sans cycle et $m = n - 1$
3. G est connexe et $m = n - 1$
4. G est connexe et supprimer une arête quelconque déconnecte G
5. G est sans cycle et ajouter une arête quelconque crée un cycle unique
6. pour tout couple de nœuds a, b de G , il existe un unique chemin de a à b

remarque : 6 \Rightarrow G est sans boucle (preuve directe en prenant $a = b$)

◦ **Preuve**

1 \Rightarrow 2 :

Par récurrence sur n : si $n = 1$ alors $m = 0$ (pas de cycle donc pas de boucle). Supposons que si G_{n-1} , connexe et sans cycle à $n - 1$ nœuds alors $m = n - 2$.

Soit G_n à n nœuds ($n > 1$), connexe et sans cycle.

G_n a au moins un nœud de degré 1 :

Il n'y a pas de nœud de degré 0 (sinon, non-connexe).

Supposons tous les degrés soient ≥ 2 , alors pour tout nœud u on peut prolonger un parcours incident en u par une arête différente de l'arête entrante.

Partant de l'arête $x_1 \rightarrow x_2$, alors il existe une arête distincte $x_2 \rightarrow u_1$, soit $u_1 = x_1$ et il existe un cycle, soit $u_1 = x_3 \neq x_1$.

Pour x_3 : il existe une arête $x_3 \rightarrow u_2$ distincte de $x_2 \rightarrow x_3$ et soit $u_2 = x_1$ ou x_2 et il existe un cycle, soit on itère le raisonnement.

Le nombre de nœuds étant fini, on aura nécessairement un i tel que u_i figure déjà dans le parcours, donc un cycle. Il y a contradiction, il existe donc un nœud de degré 1.

retour partie principale :

Soit $v \in G_n$ un nœud de degré 1, $G_n - v$ est connexe et sans-cycle, de plus il comporte $n - 1$ nœuds donc $n - 2$ arêtes, donc G_n a $n - 1$ arêtes.

2 \Rightarrow 3 :

Supposons G sans cycle et non-connexe, ses composantes connexes sont connexes et sans cycle, donc comme $1 \Rightarrow 2$, pour chaque composante connexe i , on a $m_i = n_i - 1$.

Or $m = \sum_i m_i = \sum_i (n_i - 1) = \sum_i n_i - (\# \text{ compo connexes}) = n - (\# \text{ compo connexes}) < n - 1$.

Comme par hypothèse $m = n - 1$, il y a contradiction.

3 \Rightarrow 4 :

Soit e une arête quelconque de G connexe. Supposons que $G - e$ soit connexe, il existe donc un arbre sous-tendant T de $G - e$.

Or par $1 \Rightarrow 2$, T a $n - 1$ arêtes, donc $G - e$ a plus que $n - 1$ arêtes.

Mais G a $n - 1$ arêtes donc $G - e$ a $n - 2$ arêtes, il y a contradiction.

4 \Rightarrow 5 :

G est connexe, on peut en extraire un arbre T sous-tendant (donc connexe), mais comme retirer une arête à G le déconnecte, T ne peut être que G . Donc G est un arbre et sans cycle.

Soit u et v deux nœuds quelconques de G , comme G est connexe, il existe un chemin $u \rightarrow v : u, x_1, x_2, \dots, v$. On ajoute une arête entre u et v , le parcours fermé u, x_1, \dots, v, u est un cycle.

unicité :

Supposons que l'ajout de e ait créé deux cycles C_1, C_2 . Comme G est sans cycle, e appartient à $C_1 = u, x_1, \dots, v, u$ et $C_2 = u, y_1, \dots, v, u$.

$C_1 \cup C_2 - e$ est contenue dans G et contient les deux parcours distincts : $P_x = u, x_1, \dots, v$ et $P_y = u, y_1, \dots, v$. Il existe alors x_i, \dots, x_j et y_k, \dots, y_l des parties de P_x et P_y telles que $x_i = y_k$ et $x_j = y_l$, et n'ayant aucune arête commune.

Alors $x_i, \dots, x_j, y_{l-1}, \dots, y_k$ est un cycle de G , il y a contradiction.

5 \Rightarrow 6 :

Supposons qu'il existe deux chemins distincts u, x_1, \dots, v et u, y_1, \dots, v entre les nœuds u et v de G , si on ajoute e entre u et v , on crée 2 cycles distincts : u, x_1, \dots, v, u et u, y_1, \dots, v, u , il y a contradiction.

6 \Rightarrow 1 :

G est trivialement connexe.

Supposons qu'il existe un cycle $C = u, x_1, x_2, \dots, x_n, v, u$ dans G , alors u, v et $u, x_1, x_2, \dots, x_n, v$ sont deux chemins de u à v et il y a contradiction.

4.3 Nombre d'arbres sous-tendants

* Définition

Soit un graphe G contenant une arête e , $G \cdot e$ désigne le graphe obtenu en contractant e (c'est-à-dire en supprimant e et fusionnant ses deux extrémités).

• **Théorème**

Formule de Cayley : Soit $T(G)$ le nombre d'arbres sous-tendants de G et e une arête de G qui n'est pas une boucle. Alors :

$$T(G) = T(G - e) + T(G \cdot e)$$

◦ **Preuve**

Soit T un arbre sous-tendant de G , alors soit $T \in \mathcal{A}$, l'ensemble des arbres sous-tendant G contenant e , soit $T \in \mathcal{B} = \overline{\mathcal{A}}$, l'ensemble des arbres sous-tendant G ne contenant pas e .

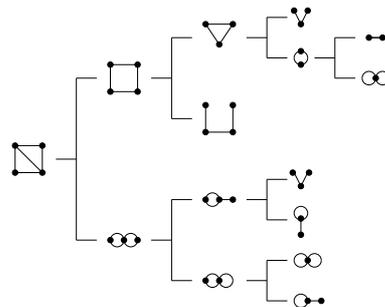
On construit une bijection entre \mathcal{A} et l'ensemble des arbres sous-tendants de $G \cdot e$, un arbre couvrant T de \mathcal{A} est associé à $T - e$ couvrant pour $G \cdot e$, ex :



On construit une bijection entre \mathcal{B} et l'ensemble des arbres sous-tendants de $G - e$, celle-ci est l'identité puisqu'un arbre couvrant de \mathcal{B} est couvrant pour $G - e$, ex :



exemple: Utilisation pour dénombrer récursivement le nombre d'arbres sous-tendants un graphe, en éliminant les cycles (on peut supprimer les boucles, qui ne peuvent pas appartenir à un arbre) :



Notons qu'un nœud est un arbre trivial. Ce graphe admet donc 8 arbres couvrants (**attention :** les arbres dénombrés admettent un unique arbre couvrant, mais ne sont pas les arbres couvrants de G .)

* **Définition**

K_n désigne le **graphe complet** à n nœuds, c'est-à-dire le graphe à n nœuds dont tous les nœuds sont connectés les uns aux autres.

• **Théorème**

Le nombre d'arbres sous-tendants de K_n est n^{n-2} .

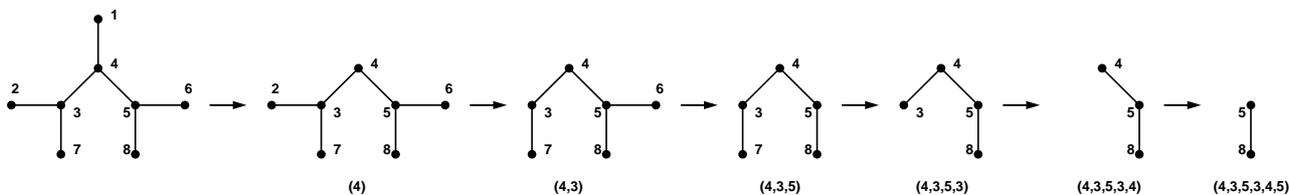
Ce théorème est également dû à Cayley, en 1889, même si la formule avait été proposée 30 ans plus tôt par le mathématicien russe Borchardt. Entre autres choses, Cayley est également connu pour l'invention du calcul matriciel.

o **Preuve**

Soit l'ensemble des sommets de $K_n : V = \{1, 2, \dots, n\}$. n^{n-2} est le nombre de séquences de longueur $n-2$ que l'on peut former à partir de V . On va démontrer la correspondance entre l'ensemble des arbres sous-tendant de K_n et l'ensemble de ces séquences.

Commençons par construire une méthode qui, à partir d'un arbre, lui fait correspondre une et une seule séquence. A chaque arbre T , on associe une séquence $(t_1, t_2, \dots, t_{n-2})$ comme suit. V est un ensemble ordonné et soit s_1 le premier sommet de degré 1 dans T . Le voisin unique de s_1 est t_1 . On retire s_1 et on appelle s_2 le plus petit sommet de degré 1 dans $T - s_1$. Le voisin de s_2 est t_2 . On procède jusqu'à ce que t_{n-2} soit défini, c'est-à-dire qu'il ne reste plus qu'un lien entre deux sommets.

exemple:



Notons que les éléments apparaissant dans la séquence sont les sommets de degré $d > 1$ et ils apparaissent $d-1$ fois. De plus, on peut voir que différents arbres sous-tendant de K_n génèrent des séquences différentes.

A partir de chaque séquence, on peut reconstruire l'arbre sans ambiguïté en suivant l'algorithme suivant.

- on initialise l'ensemble S à l'ensemble des sommets du graphe
- on prend le sommet $s_i \in S$ de plus petit indice n'apparaissant pas dans la séquence, et on le lie au premier élément de la séquence
- on retire le premier élément de la séquence, on retire s_i à S et on itère le raisonnement jusqu'à épuisement de la séquence
- au terme de la procédure, on ajoute un lien entre les deux éléments restants de S

A noter que cette dernière étape se fait sans information supplémentaire, raison pour laquelle elle n'a pas besoin d'être stockée dans la séquence.

remarque : cette séquence est aussi une méthode économique de stockage d'un graphe

En repartant de l'exemple précédent, le décodage du graphe de 8 sommets à partir de la séquence (4, 3, 5, 3, 4, 5) donne :

- Premier élément de N pas dans la séquence (4, 3, 5, 3, 4, 5) : 1 \rightarrow arête entre 1 et 4.
- Premier élément de $N \setminus \{1\}$ pas dans la séquence (3, 5, 3, 4, 5) : 2 \rightarrow arête entre 2 et 3.
- Premier élément de $N \setminus \{1, 2\}$ pas dans la séquence (5, 3, 4, 5) : 6 \rightarrow arête entre 6 et 5.
- Premier élément de $N \setminus \{1, 2, 6\}$ pas dans la séquence (3, 4, 5) : 7 \rightarrow arête entre 7 et 3.
- Premier élément de $N \setminus \{1, 2, 6, 7\}$ pas dans la séquence (4, 5) : 3 \rightarrow arête entre 3 et 4.
- Premier élément de $N \setminus \{1, 2, 3, 6, 7\}$ pas dans la séquence (5) : 4 \rightarrow arête entre 4 et 5.
- Après ces $n - 2 = 6$ étapes, on connecte les deux derniers éléments dans $N \setminus \{1, 2, 3, 4, 6, 7\}$: 5 et 8.

Dans le cas du graphe complet, tous les nœuds sont connectés, donc toute séquence construite comme décrit

précédemment est un arbre couvrant de K_n . On a donc construit une bijection entre l'ensemble des arbres et l'ensemble des séquences. Pour chacun des $n - 2$ éléments de la séquence, il y a n choix donc $T(K_n) = n^{n-2}$.

A noter que n^{n-2} est le nombre d'arbres sous-tendant de K_n , mais que le nombre d'arbres non-isomorphes est nettement plus petit. Par exemple, pour K_6 , on en a 6, à comparer avec $6^4 = 1296$.

4.4 Problème de l'arbre sous-tendant de poids minimum

Revenons-en au problème qui a motivé ce chapitre : celui de l'effeuillage du réseau de la SNCB tout en préservant sa connexité. Soit w_{ij} le coût d'entretien des rails entre les gares i et j . On tente donc de trouver dans ce graphe pondéré un sous-graphe connexe sous-tendant de poids minimum. Etant donné que les poids sont positifs, on peut en déduire que la solution est un arbre sous-tendant. La question devient : comment, dans la pratique, trouver cet arbre sous-tendant de poids minimum ?

Algorithme naïf (force brute) :

On teste tous les arbres possibles et on conserve celui de poids minimum. Pour n nœuds, le coût est de n^{n-2} , pire qu'exponentiel...

Considérons d'abord le cas trivial où $w(e) = 1$ est $\forall e \in E$: chaque arête a le même poids. Dans un cas tel que celui-ci, tous les arbres sous-tendant ont le même poids. Le problème revient donc à la construction d'un tel arbre. L'algorithme suivant permet de le faire :

1. choisir une arête e_1
2. si les arêtes e_1, \dots, e_i ont été choisies, choisir l'arête e_{i+1} dans $E \setminus \{e_1, \dots, e_i\}$ de telle sorte que le graphe formé par $\{e_1, \dots, e_i, e_{i+1}\}$, noté $G[\{e_1, \dots, e_i, e_{i+1}\}]$ est acyclique
3. arrêter l'étape 2 quand $n - 1$ arêtes ont été sélectionnées.

Cet algorithme génère clairement un arbre. De plus, il est clairement sous-tendant si le graphe initial est connexe. Il a été étendu par Kruskal, en 1956, pour résoudre le problème général de graphe avec des poids positifs.

Algorithme de Kruskal (1956) :

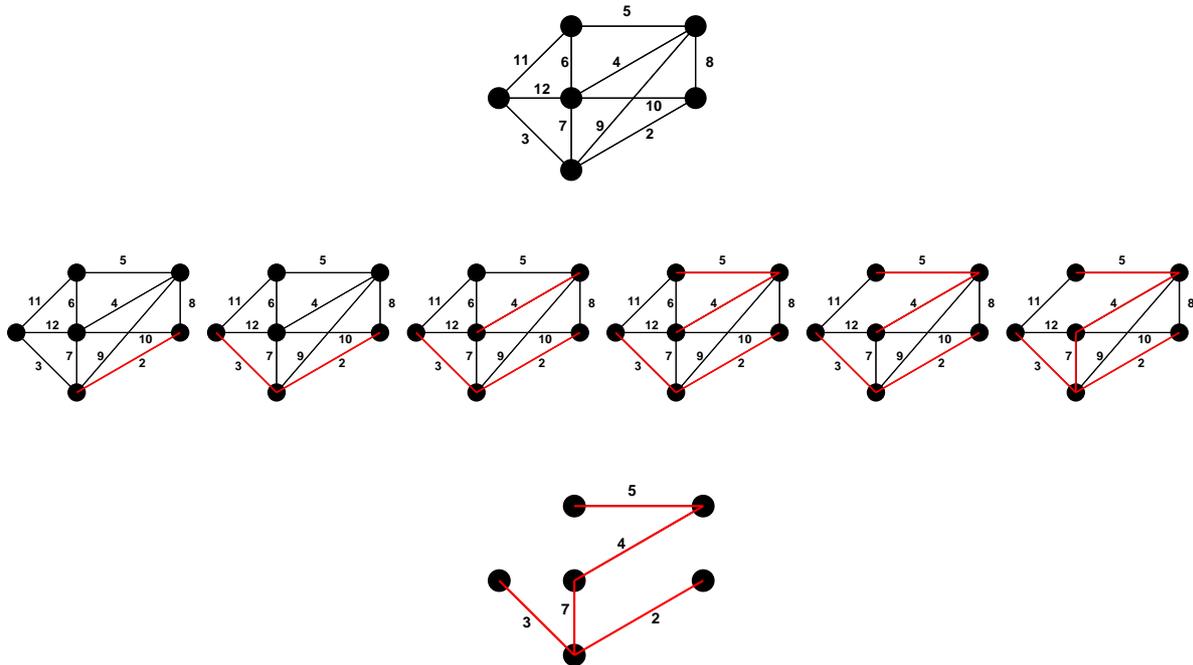
Soit G un graphe pondéré à n nœuds, soit T un ensemble d'arêtes initialement vide.

- on trie les arêtes par poids croissant, on obtient la séquence S
- tant que $|T| < n - 1$:
 - on considère le premier élément e de S (élément de moindre poids) et on le retire de S
 - si $T \cup \{e\}$ est sans cycle, alors $T \leftarrow T \cup \{e\}$

Au terme du processus, le graphe formé des arêtes de T est un arbre sous-tendant de poids minimum.

remarque : un tel algorithme est dit **vorace** (ou glouton), il choisit à chaque étape la meilleure amélioration possible à cette étape

exemple: On cherche l'arbre sous-tendant de poids minimum de :



• **Théorème**

L'algorithme de Kruskal est **correct** : il produit un arbre sous-tendant de poids minimum.

○ **Preuve**

Soit T^* le graphe produit par l'algorithme, il a $n - 1$ arêtes et n'a pas de cycle, c'est donc un arbre (c.f. Théorème 4.2).

Soit T un arbre sous-tendant de poids minimum, d'arêtes triées par ordre de poids croissant : e_1, e_2, \dots, e_{n-1} : $w(e_1) \leq w(e_2) \leq \dots \leq w(e_{n-1})$.

On procède par l'absurde : on suppose que $T \neq T^*$.

Soit f la fonction associant à un arbre A le plus petit indice a tel que $e_a \notin A$ avec $e_a \in T^*$, on note $f(T) = i$. Autrement dit : e_1, e_2, \dots, e_{i-1} sont dans T et T^* , mais $e_i \notin T$.

De plus on choisit T parmi les arbres couvrants de poids minimum tel que $f(T)$ soit maximal.

$T \cup \{e_i\}$ contient un unique cycle C (Th. 4.2).

T^* est un arbre donc sans cycle, il existe donc au moins une arête e_j de C qui n'est pas dans T^* . De plus, par construction, on a $j > i$

$T' = T \cup \{e_i\} - e_j$ compte $n - 1$ arêtes et pas de cycle, c'est donc un arbre et il est couvrant.

Son poids est de : $w(T') = w(T) + w(e_i) - w(e_j)$.

De plus si on suit l'algorithme de Kruskal, arrivé à la séquence e_1, e_2, \dots, e_{i-1} ; l'ajout de e_i ne crée pas de cycle (ces arêtes sont toutes dans T'), l'ajout de e_j non plus (ces arêtes sont toutes dans T).

Si l'algorithme ajoute e_i , c'est que $w(e_i) \leq w(e_j)$, et donc $w(T') \leq w(T)$. Comme T est de poids minimum, on a $w(T') = w(T)$.

Cependant, $f(T') > i = f(T)$, car :

$$T^* = e_1 e_2 \dots e_{i-1} e_i \dots$$

$$T = e_1 e_2 \dots e_{i-1} f_i \dots$$

$$T' = e_1 e_2 \dots e_{i-1} e_i \dots \quad (4.1)$$

C'est contradictoire (on a choisi $f(T)$ maximal) donc $T = T^*$.

remarque : l'algorithme de Kruskal peut être implémenté avec une complexité $O(m \cdot \log(m))$ pour un graphe à m arêtes (par exemple, tri des liens par tri fusion et vérification qu'il n'y a pas création de boucle en maintenant à jour les composantes connexes par *union find*).

4.5 Connectivité

Un mouvement terroriste aimerait mettre le monde à terre en s'attaquant à son réseau de communication. L'état aimerait enrayer une épidémie en vaccinant un nombre minimum de personnes. Deux facettes d'un même problème : Quelle serait sa meilleure stratégie ? Quelles sommets, ou arêtes retirer du graphe ?

* *Définition*

Pour un graphe connexe, une **coupe de sommets** est un ensemble de sommets qui déconnecte le graphe quand on l'en retire.

* *Définition*

Pour un graphe connexe, une **coupe d'arêtes** est un ensemble d'arêtes qui déconnecte le graphe quand on l'en retire.

* *Définition*

Un graphe est dit **k -connexe** si retirer $k - 1$ nœuds quelconques laisse le graphe connexe. Autrement dit, un graphe est **k -connexe** si toutes les coupes de sommets sont au moins de taille k .

* *Définition*

On appelle **connectivité** d'un graphe la taille de la plus petite coupe de sommets. La connectivité est égale à la valeur de k la plus grande pour laquelle le graphe est k -connexe.

remarque : par définition, on dit que pour un graphe complet à n nœuds, elle est de $n - 1$

* *Définition*

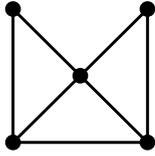
Un graphe est dit **k -arête connexe** si retirer $k - 1$ arêtes quelconques laisse le graphe connexe. Autrement dit, un graphe est **k -arête-connexe** si toutes les coupes d'arêtes sont au moins de taille k .

* *Définition*

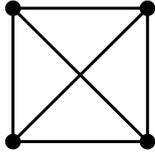
On appelle **arête-connectivité** d'un graphe la taille de la plus petite coupe d'arêtes.

remarque : un graphe connexe est au moins 1-arête-connexe et 1-connexe

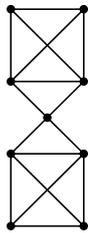
exemple:



- G est 2-connexe mais pas 3-connexe. Connectivité : 2
- G est 2-arête-connexe mais pas 3-arête-connexe. Arête-connectivité : 2



- K_4 est 3-connexe, par convention sa connectivité est 3
 - K_4 est 3-arête-connexe mais pas 4-arête-connexe
- remarque :** plus généralement K_n est de connectivité et d'arête-connectivité $n - 1$



- G est 1-connexe mais pas 2-connexe
- G est 2-arête-connexe mais pas 3-arête-connexe
- G est de degré minimum 3

• **Théorème**

connectivité \leq arête-connectivité \leq degré minimum

◦ **Preuve**

arête-connectivité \leq degré minimum :

Si le degré minimum est de δ , on trouve trivialement une coupe à δ arêtes : celle qui isole le nœud de degré minimum.

connectivité \leq arête-connectivité :

On note la connectivité k , l'arête-connectivité k' et on procède par récurrence sur k' :

Si $k' = 0$ le graphe est déconnecté donc $k = 0$. Supposons que connectivité \leq arête-connectivité pour tout graphe d'arête-connectivité $< k'$.

Soit G d'arête connectivité k' , il admet une coupe de k' arêtes, soit e une de ces arêtes.

$G - e$ est d'arête-connectivité $k' - 1$, donc par hypothèse $k_{G-e} \leq k' - 1$.

Soit S une coupe de nœuds de taille minimum de $(G - e)$, donc $G - S - e$ est déconnecté. Par hypothèse $|S| \leq k' - 1$. Est-ce que $G - S$ est déconnecté ?

- si oui, alors S est une coupe de nœuds de G , et $k_G \leq |S| \leq k' - 1 \leq k'$

- si non, $G - S$ est connexe et $\{e\}$ est une coupe d'arêtes pour $G - S$.

$e = (u, v)$, si on retire u , on retire e et on déconnecte $G - S$ donc $G - (S \cup \{u\})$ est déconnecté.

$S \cup \{u\}$ est une coupe de nœuds pour G de taille $|S| + 1 \leq k' - 1 + 1 = k'$

Dans les deux cas il existe une coupe de nœuds de taille $\leq k'$.

4.6 Théorème de Menger

• **Théorème**

Cas 2-connexe : Un graphe à au moins 3 nœuds est 2-connexe \Leftrightarrow toute paire de nœuds distincts est reliée

par au moins 2 chemins dont les nœuds internes sont tous distincts.

Corrolaire : dans un graphe 2-connexe, 2 nœuds sont toujours sur un même cycle.

Théorème de Menger (1927) : (généralisation) Un graphe à au moins $(k + 1)$ nœuds est k -connexe \Leftrightarrow toute paire de nœuds distincts est reliée par au moins k chemins dont les nœuds internes sont tous distincts. Autrement dit, si $n \geq 3$ alors k -connexe $\Leftrightarrow \exists k$ chemins disjoints de u à v .

◦ **Preuve**

preuve du cas 2-connexe

1 \Leftrightarrow 2 :

$\forall u, v$, il existe au moins 2 chemins de u à v dont les nœuds internes sont distincts : u, x_1, \dots, x_n, v et u, y_1, \dots, y_m, v . Si on supprime un des x_i , il reste le chemin u, y_1, \dots, y_m, v , le graphe est connexe donc 2-connexe.

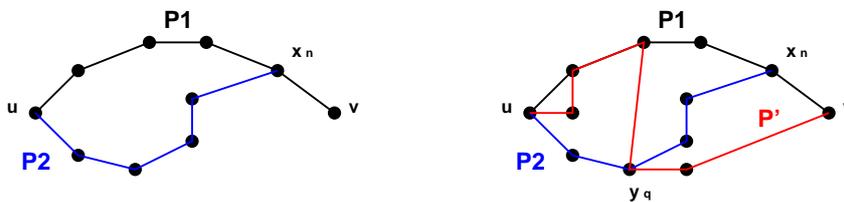
1 \Rightarrow 2 :

On procède par récurrence sur la distance $d(u, v)$ entre u et v . Si $d(u, v) = 1$ et que le graphe est 2-connexe, il est donc au moins 2-arête-connexe, donc si on retire l'arête (u, v) , il doit exister un autre chemin entre u et v . Supposons que pour toutes les paires à distance $< d(u, v)$, il existe au moins 2 chemins pour aller d'un nœud à l'autre.

Considérons un plus court chemin de u à v : $P_1 = u, x_1, \dots, x_n, v$. il existe donc le chemin u, x_1, \dots, x_n pour aller de u à x_n , et par hypothèse de récurrence, il en existe un second dont les nœuds internes sont distincts : $P_2 = u, z_1, \dots, z_p, x_n$.

Si on supprime x_n , la 2-connexité assure qu'il existe un second chemin de u à v : $P' = u, y_1, \dots, y_m, v$. Soit y_q le dernier nœud commun à P' et $P_1 \cup P_2$,

- soit $y_q \in P_2$ alors u, x_1, \dots, x_n, v et $u, z_1, \dots, y_q, \dots, y_m, v$
 - soit $y_q \in P_1$ alors $u, x_1, \dots, y_q, \dots, y_m, v$ et $u, z_1, \dots, z_p, x_n, v$
- ... sont 2 chemins distincts de u à v .



4.7 Un réseau de communication fiable

Problème : comment construire un réseau de communication avec le moins d'arêtes possibles et qui reste connexe après la suppression de $k - 1$ nœuds ?

• **Théorème**

Tout graphe k -connexe à n nœuds possède au moins $\frac{kn}{2}$ arêtes.

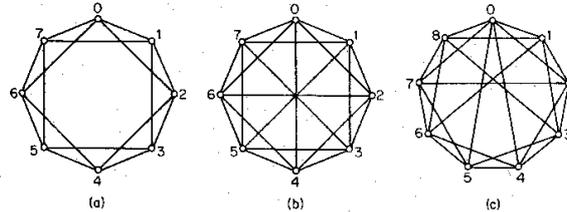
◦ **Preuve**

Comme le degré minimum est supérieur à la connexité, pour chaque nœud i , son degré $\delta_i \geq k$.
 Donc $\sum_i \delta_i \geq k.n$, or $\sum_i \delta_i = 2m$ avec m le nombre total d'arêtes, donc $m \geq \frac{kn}{2}$.

• **Théorème**

On peut construire des graphes k -connexes qui ont exactement $\frac{kn}{2}$ (si kn pair) ou $\left\lceil \frac{kn}{2} \right\rceil + 1$ arêtes (si kn impair), ce sont **les graphes d'Harary** $H_{k,n}$ (k -connexité, n nœuds).

- **si k pair**, soient les nœuds v_0, v_1, \dots, v_{n-1} , on relie i et j si :
 $i - \frac{k}{2} \leq j \leq i + \frac{k}{2}$ (modulo n).
- **si k impair, n pair**, soient les nœuds v_0, v_1, \dots, v_{n-1} , on relie i et j si :
 $i - \frac{k-1}{2} \leq j \leq i + \frac{k-1}{2}$ (modulo n),
 ou si $i \in [1; \frac{n}{2}]$ et $j = i + \frac{n}{2}$.
- **si k impair, n impair**, soient les nœuds v_0, v_1, \dots, v_{n-1} , pour $i \neq 0$ on relie i et j si :
 $i - \frac{k-1}{2} \leq j \leq i + \frac{k-1}{2}$ (modulo n),
 ou si $i \in [1; \frac{n-1}{2}]$ et $j = i + \frac{n+1}{2}$,
 ou si $i = 0$ et $j = \frac{n-1}{2}$ ou $j = \frac{n+1}{2}$.



Graphes d'Harary : (a) $H_{4,8}$, (b) $H_{5,8}$, (c) $H_{5,9}$. Extrait de Bondy & Murty.

4.8 Notion de robustesse

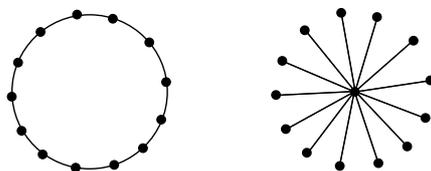
4.8.1 Robustesse des réseaux

La définition précise de la robustesse d'un système dépend du contexte, une formulation générique serait : *la capacité d'un système à assurer des performances stables sur une large gamme de paramètres*. Nous considérons ici le cas de réseaux, c'est-à-dire d'objets physiques ou abstraits qui peuvent être représentés mathématiquement par des graphes (réseaux d'amitié, sociaux, de communication, financiers, ...). Dans ce cas la question de la robustesse est souvent liée à la capacité du graphe à rester connexe en cas de suppression d'arêtes ou de sommets.

Prenons l'exemple du protocole de suppression suivant :

- on choisit aléatoirement un sommet du graphe et on le retire,
- on itère cette procédure jusqu'à obtenir plusieurs composantes connexes,

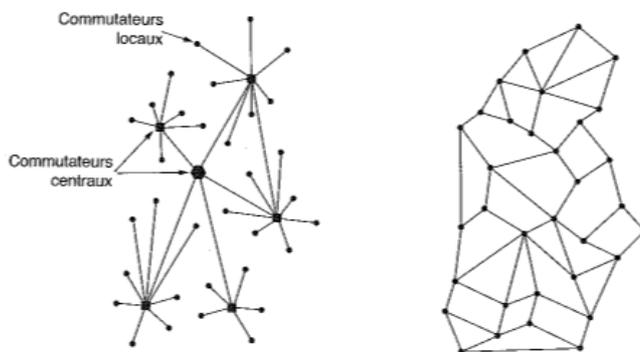
– on compte le nombre de suppressions pour déconnecter le graphe.
Puis considérons alors les deux réseaux suivants :



Pour le premier, deux suppressions sont (en général) suffisantes pour déconnecter le graphe, il est peu robuste à ce protocole ; en revanche, une attaque aléatoire ne déconnecte le second graphe que si le sommet central est touché, ce qui est un événement peu probable. Mais en cas d'attaque ciblée sur le réseau, le second peut être facilement déconnecté. La robustesse dépend également du type de risque considéré.

4.8.2 La robustesse d'Internet

Prenons l'exemple historique de l'ancêtre d'Internet : ARPANET. Dans les années 60, en pleine Guerre Froide, le Département de la Défense des États-Unis veut créer un réseau de communication capable de résister à une attaque nucléaire soviétique. Le modèle classique de réseau de communication à l'époque est le réseau téléphonique à commutateurs, celui-ci est hiérarchisé et vulnérable aux attaques ciblées, comme le réseau en étoile. Paul Baran propose alors une structure à commutateurs distribués, sans hiérarchie et donc beaucoup moins vulnérable aux attaques car il n'y a pas de cibles privilégiées (c.f. Figure).



Gauche : réseau téléphonique à commutateurs centralisés ; droite : réseau à commutateurs distribués de Baran. Extrait de *Réseaux* (Tanenbaum & Wetherall).

La contrepartie à ce système est la nécessité d'adapter les protocoles à ces commutateurs distribués, c'est dans ce but que seront définis les protocoles historiques de communication par commutation de paquets : TCP/IP. L'Internet actuel, s'il n'est pas dénué de hiérarchie, est hérité de cette structure décentralisée.

Chapitre 5

Graphes hamiltoniens

5.1 Chemins et cycles hamiltoniens

* *Définition*

Un **chemin est hamiltonien** s'il passe par chaque nœud du graphe une et une seule fois.

* *Définition*

Un **cycle est hamiltonien** s'il passe par chaque nœud du graphe une et une seule fois.

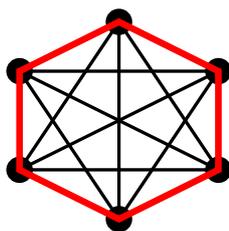
* *Définition*

Un **graphe est hamiltonien** s'il possède un cycle hamiltonien.

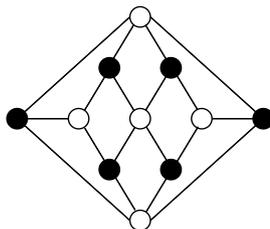
Remarque : si on ajoute des arêtes à un graphe hamiltonien, le résultat est hamiltonien, cette propriété est dite **monotone**.

exemple:

graphe complet : (toujours hamiltonien)



graphe non-hamiltonien : Ce graphe (le graphe de Herschel) est biparti avec 6 nœuds noirs et 5 blancs, un cycle doit visiter autant de nœuds noirs que blancs, il ne peut donc pas être hamiltonien.



5.2 Condition nécessaire

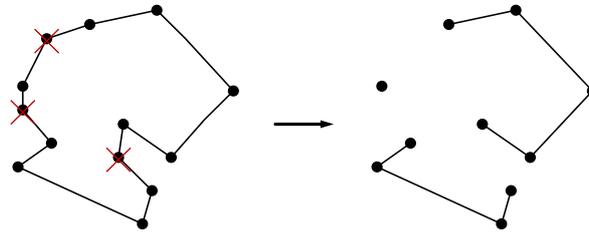
La contraposée du théorème suivant fournit une condition nécessaire pour avoir un graphe hamiltonien.

• Théorème

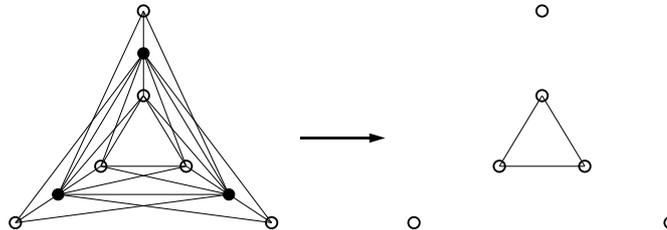
Si on retire k nœuds quelconques d'un graphe hamiltonien, on obtient au plus k composantes connexes.

◦ Preuve

Soit $v_1v_2\dots v_nv_1$ un cycle hamiltonien de G , en enlevant k nœuds v_a, v_b, \dots, v_k du cycle on crée au plus k chemins : $v_{a+1}\dots v_{b-1}$; $v_{b+1}\dots v_{c-1}$; ...; $v_{j+1}\dots v_{k-1}$; $v_{k+1}\dots v_nv_1\dots v_{a-1}$. Tous les nœuds sur un même chemin étant dans une même composante connexe, on a créé au plus k composantes connexes.



exemple: Le graphe suivant n'est pas hamiltonien car en enlevant les 3 nœuds noirs, on obtient 4 composantes connexes.



5.3 Conditions suffisantes

5.3.1 Théorème de Dirac

Le théorème suivant (Dirac, 1952) peut permettre de prouver qu'un graphe est hamiltonien sans avoir à trouver un cycle hamiltonien.

• Théorème

Un graphe simple à $n \geq 3$ nœuds tel que le degré minimum est d'au moins $n/2$ est hamiltonien.

remarque : autrement dit, les graphes suffisamment "denses" sont hamiltoniens

◦ Preuve

Supposons qu'il existe un graphe simple de degré minimum $n/2$ et non-hamiltonien, on choisit alors G maximal avec ces propriétés, c'est-à-dire que l'ajout d'une arête rend G hamiltonien (on peut le faire car l'ensemble des graphes simples à n nœuds est fini). Il existe v_1 et v_n dans G non adjacents (si tous les

sommets sont adjacents, le graphe est complet, donc hamiltonien).

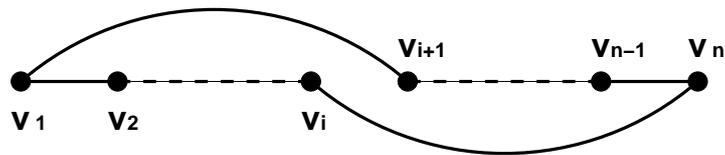
Comme G est maximal, $G + v_1v_n$ est hamiltonien. Les cycles hamiltoniens de $G + v_1v_n$ passent par v_1v_n , donc il existe un chemin hamiltonien $v_1v_2\dots v_n$ dans G .

Soient :

- $S = \{v_i : v_{i+1} \text{ est adjacent dans } G \text{ à } v_1\}$
- $T = \{v_i : v_i \text{ est adjacent dans } G \text{ à } v_n\}$

On sait que $v_n \notin T$ et $v_n \notin S$ donc $|S \cup T| < n$.

De plus, $S \cap T = \emptyset$, en effet si $\exists v_i$ tel que $v_i \in S \cap T$, alors $v_1v_2\dots v_iv_nv_{n-1}\dots v_{i+1}v_1$ serait un cycle hamiltonien (c.f. figure).



De plus, $|S \cup T| = |S| + |T| - |S \cap T|$. Or $|T| = \delta(v_n) \geq n/2$ et $|S| = \text{nombre d'antécédents dans le chemin aux voisins de } v_1 = \delta(v_1)$ donc $|S| \geq n/2$. Il y a contradiction donc G n'existe pas.

C'est une condition suffisante, mais clairement pas nécessaire : un graphe constitué d'un unique cycle passant par tous les nœuds est hamiltonien.

5.3.2 Théorème d'Ore

On peut donner une condition suffisante plus générale que le théorème de Dirac : le théorème d'Ore (1960).

• Théorème

Dans un graphe simple G à $n \geq 3$ nœuds, si pour toute paire de sommets non-adjacents (u, v) , $\delta(v) + \delta(u) \geq n$ alors G est hamiltonien.

La démonstration est exactement du même type que celle du théorème de Dirac.

5.4 Graphes eulériens, graphes hamiltoniens et problèmes d'algorithmique

Nous avons vu une condition nécessaire et suffisante pour déterminer "facilement" si un graphe est eulérien (le théorème d'Euler). En revanche, s'il existe des conditions nécessaires et des conditions suffisantes pour prouver qu'un graphe est hamiltonien, **on ne dispose d'aucune condition nécessaire et suffisante simple**.

Du point de vue algorithmique, on dispose d'un algorithme efficace (c'est à dire de complexité temporelle polynomiale) pour vérifier qu'un graphe est eulérien, par exemple : vérifier que tous les sommets sont de degré pair, mais on n'a pas de tel algorithme pour vérifier qu'un graphe est hamiltonien.

5.4.1 Graphes hamiltoniens et NP-complétude

Trouver un tel algorithme est un des problèmes les plus importants des mathématiques et de l'informatique actuelle. En effet, le problème de décision : *un graphe donné comporte-t-il un cycle hamiltonien* est un problème **NP-complet**. Un problème de décision est dit NP-complet si :

- a) il appartient à la classe de complexité NP (*non-deterministic polynomial time*) : pour toute instance vraie, on peut vérifier en temps polynomial que cette instance est effectivement vraie,
- b) tout problème de NP peut être réduit en temps polynomial à ce problème (il est donc au moins aussi difficile que tout problème de NP).

À l'heure actuelle, on sait donc vérifier en temps polynomial qu'un cycle est hamiltonien, mais on ne sait pas trouver en temps polynomial un tel cycle.

Trouver un tel algorithme, c'est montrer que les problèmes les plus difficiles de la classe NP peuvent être résolus en temps polynomial, autrement dit c'est montrer que $P=NP$. L'Institut Clay récompense d'un million de dollars celui qui trouvera un tel algorithme ou montrera qu'un tel algorithme n'existe pas (problèmes du Prix du Millénaire).

5.4.2 D'autres problèmes classiques sur les parcours de graphes (pondérés)

Problème du postier chinois : (*route inspection problem*)

Un postier veut passer dans chaque rue au moins une fois, de façon à parcourir la plus petite longueur possible. Dans un graphe pondéré, trouver le parcours fermé le plus court qui passe par toutes les arêtes au moins une fois. Meigu Guan a trouvé un algorithme efficace de résolution du problème en 1962.

Notons que si le graphe comporte un cycle eulérien, celui-ci est la solution optimale.

Problème du voyageur de commerce : (*travelling salesman problem*)

Un voyageur de commerce doit passer par chaque ville d'un parcours exactement une fois et revenir à son point de départ en parcourant la plus petite distance possible. Étant donné un graphe complet pondéré dont les poids des arêtes sont les distances entre les sommets, trouver le parcours fermé de longueur totale minimale passant exactement une fois par chaque sommet. La résolution par la force brute de ce problème (énumérer toutes les solutions possibles) soulève le problème de l'explosion combinatoire du nombre de cycles hamiltoniens d'un graphe simple complet à n sommets (qui évolue en $n!$). La solution par une recherche locale (aller au plus proche voisin encore non-visité) n'est pas non plus optimale en général. On ne connaît pas d'algorithme efficace pour résoudre ce problème. Le problème décisionnel associé (*pour une distance D , existe-t-il un chemin plus court passant par tous les sommets du graphe ?*) est NP-complet. Ce problème est beaucoup utilisé comme terrain d'exploration de métaheuristiques d'optimisation : algorithmes génétiques, de colonies de fourmis...

Chapitre 6

Mariages, couplages et couvertures

Dans tout ce chapitre, si les définitions peuvent s'appliquer à tout type de graphes, les démonstrations se limitent aux **graphes connexes**. Pour des graphes non-connexes, il suffirait de considérer chacune des composantes connexes séparément et d'y appliquer les théorèmes démontrés.

6.1 Définitions de base

* *Définition*

Un **couplage** dans un graphe est un ensemble M d'arêtes tel que M ne contient pas de boucles et deux arêtes quelconques de M n'ont pas d'extrémité commune.

* *Définition*

Un **couplage maximum** est un couplage dont le nombre d'arêtes est maximal.

* *Définition*

Un **couplage parfait** est un couplage qui est incident à tous les nœuds.

Remarques :

- un couplage parfait est maximum
- un graphe qui admet un couplage parfait a un nombre de nœuds pair

6.2 Couplages maximaux, chemins alternés et chemins augmentés

* *Définition*

Soit un couplage M , un **chemin M -alterné** est un chemin qui passe par une arête de M et une arête hors de M .

* *Définition*

Soit un couplage M , un **chemin M -augmenté** est un chemin M -alterné dont les nœuds d'origine et de destination sont distincts et ne sont pas incidents à une arête de M .

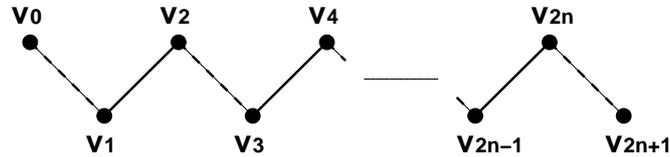
• **Théorème**

Théorème de Berge (1957) : Un couplage M est maximum si et seulement si il n'y a pas de chemin M -augmenté.

◦ **Preuve**

1 ⇒ 2 :

M est maximum et compte n arêtes, supposons qu'il existe un chemin M -augmenté : $v_0v_1\dots v_{2n+1}$.



Soit $M' = \{v_0v_1, v_2v_3, \dots, v_{2n}v_{2n+1}\}$. v_0 et v_{2n+1} n'étant par définition, incidents à aucune arête de M , ils sont incidents à une et une seule arête de M' . De plus, les nœuds internes sont par construction incidents à une et une seule arête de M' donc M' est un couplage. Ce couplage compte $n + 1$ arêtes, M n'est donc pas maximum, il y a contradiction.

2 ⇒ 1 :

On démontre la contraposée **Non(1) ⇒ Non(2)**.

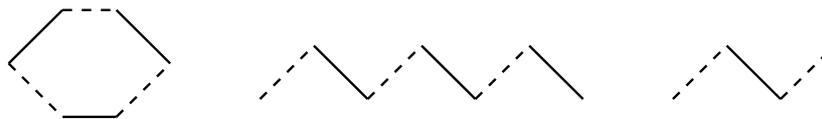
Supposons M non maximum, soit M^* un couplage maximum et H l'ensemble des arêtes appartenant soit à M soit à M^* mais pas aux deux, autrement dit :

$$H = (M \cup M^*) \setminus (M \cap M^*)$$

Soit G le graphe constitué des arêtes H et des nœuds adjacents, G ne contient donc que des nœuds de degré 1 ou 2. En effet, si un nœud est de degré 3 ou plus alors il serait adjacent à au moins 2 arêtes soit dans M soit dans M^* , donc soit M soit M^* ne serait pas un couplage.

Les composantes connexes de G sont donc :

- soit des cycles de longueurs paires alternant entre arête de M et de M^* (donc même nombre dans chacun des deux ensembles),
- soit des chemins alternant entre arête de M et de M^* .



Comme $|M^*| > |M|$, il y a dans H au moins une composante connexe qui est un chemin commençant et terminant par une arête de M^* , par définition ce chemin est M -augmenté (c'est-à-dire **Non(2)**).

6.3 Couplage de graphes bipartis

6.3.1 Théorème de Hall et conséquences

• Théorème

Théorème de Hall ou du mariage (1935) : Un graphe biparti avec bipartition (X, Y) possède un couplage incident à tous les nœuds de X si et seulement si $\forall S \subseteq X$, le nombre de voisins de S est au moins $|S|$, (on notera $|\mathcal{V}(S)| \geq |S|$).

◦ Preuve

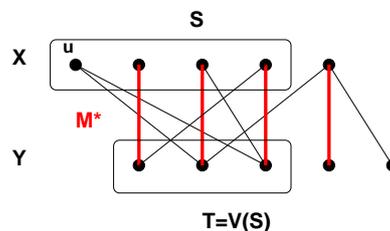
1 \Rightarrow 2 :

$\forall S \subseteq X$, $\mathcal{V}(S)$ contient les nœuds couplés à chacun des éléments de S par le couplage, qui sont au nombre de $|S|$ donc $|\mathcal{V}(S)| \geq |S|$.

2 \Rightarrow 1 :

On procède par l'absurde : supposons que (X, Y) est un graphe biparti tel que $\forall S \subseteq X$, $|\mathcal{V}(S)| \geq |S|$ mais ne contenant aucun couplage incident à tous les nœuds de X . Soit M^* un couplage maximum de (X, Y) , M^* n'est donc pas incident à tous les nœuds de X . Il existe donc u , un sommet de X non incident à M^* .

On appelle Z l'ensemble des nœuds que l'on peut atteindre par un chemin M^* -alterné partant de u . On appelle $S = Z \cap X$ et $T = Z \cap Y$. Notons que u est le seul sommet de S non incident à M^* , car si il en existe un autre v , cela implique qu'on peut construire un chemin M^* -alterné liant u à v qui serait M^* -augmenté, ce qui est contradictoire avec M^* maximum.



Les sommets de $S \setminus \{u\}$ sont couplés par M^* aux sommets de T , donc $|S| - 1 = |T|$.

De plus, $T \subseteq \mathcal{V}(S)$ car tout élément de T doit être lié à un élément de S pour appartenir au chemin M^* -alterné.

Par ailleurs, pour tout élément y de $\mathcal{V}(S)$, il existe un lien (x, y) le liant à $x \in S$. Observons d'abord que lorsqu'on construit des chemins M^* -alternés depuis u on découvre les nœuds de X par des arêtes de M^* et les nœuds de Y par des arêtes hors de M^* . Alors :

- soit $(x, y) \in M^*$, comme on doit découvrir x par une arête du couplage, on est passé par y pour découvrir x et y appartient nécessairement à T ,
- soit $(x, y) \notin M^*$, donc il existe une arête de M^* par laquelle on est arrivé en x et on peut donc construire un chemin M^* -alterné qui arrive en y , autrement dit $y \in T$.

Donc $\forall y \in \mathcal{V}(S), y \in T$, c'est-à-dire $\mathcal{V}(S) \subseteq T$.

Finalement, $\mathcal{V}(S) = T$ donc $|\mathcal{V}(S)| = |T| = |S| - 1 < |S|$, il y a contradiction et \exists au moins un couplage incident à tous les nœuds de X .

* Définition

Un graphe est *k-régulier* si tous les nœuds sont de degré k .

Corollaire du théorème de Hall : Tout graphe biparti k -régulier possède un couplage parfait.

o Preuve

Soit (X, Y) la bipartition. Le graphe doit compter $k|X| = k|Y|$ arêtes, donc $|X| = |Y|$.

Soit $S \subseteq X$, E_1 l'ensemble des arêtes incidentes à S et E_2 les arêtes incidentes à $\mathcal{V}(S)$.

$E_1 \subseteq E_2$, or $|E_1| = k|S|$ et $|E_2| = k|\mathcal{V}(S)|$ donc $|\mathcal{V}(S)| \geq |S|$. D'après le théorème du mariage, il y a un couplage incident à tous les nœuds de X , et comme $|X| = |Y|$, il est également incident à tous les nœuds de Y , c'est donc un couplage parfait.

6.3.2 Couvertures

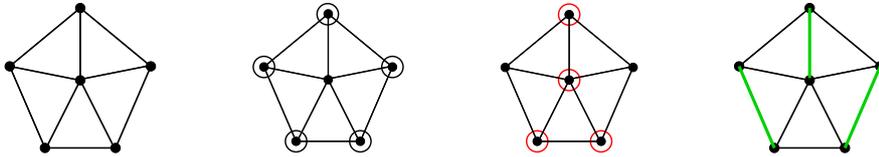
* Définition

Une **couverture de sommets** d'un graphe est un ensemble de sommets incident à toutes les arêtes.

* Définition

Une **couverture de sommets minimum** d'un graphe est une couverture de sommets avec un nombre minimum de sommets.

exemple:



Les 5 sommets entourés du second schéma forment une couverture de sommets. Les 4 sommets entourés en rouge du troisième schéma forment une couverture minimum. Les 3 arêtes vertes du quatrième schéma forment un couplage maximum du graphe.

• Théorème

Si K est une couverture de sommets et M est un couplage, alors $|M| \leq |K|$. C'est vrai en particulier pour un couplage maximum et pour une couverture de sommets minimum.

o Preuve

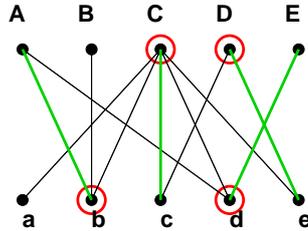
Une arête du couplage est incidente à au moins un sommet de la couverture, puisque la couverture doit être incidente à toute arête du graphe, donc du couplage. On peut donc construire une fonction injective du couplage vers la couverture (c'est-à-dire telle que toute image a au plus un antécédent).

Corollaire : Si K est une couverture de sommets, M un couplage et que $|M| = |K|$, alors K est minimum et M est maximum.

Preuve : on appelle M^* un couplage maximum, K^* une couverture minimum.

Alors $|M| \leq |M^*| \leq |K^*| \leq |K|$, donc si $|M| = |K|$, $|M| = |M^*| = |K^*| = |K|$.

exemple: Problème du site de rencontres : le site dispose d'une liste de garçons et d'une liste de filles. Chaque fille déclare si un garçon la séduit ou pas, comment effectuer l'appariement pour que chaque fille soit appariée avec un garçon qui la séduit ? Il s'agit donc de réaliser un couplage maximum dans un graphe biparti.



Les quatre nœuds entourés en rouge forment une couverture, les quatre liens en vert forment un couplage, donc il s'agit d'un couplage maximum et d'une couverture minimum.

6.3.3 Théorème de Kőnig

• **Théorème**

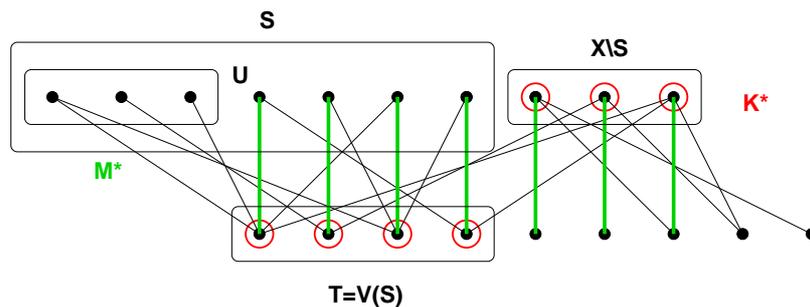
Théorème de Kőnig (1931) : Dans un graphe biparti, si K^* est une couverture de sommets minimum et M^* un couplage maximum alors $|K^*| = |M^*|$.

◦ **Preuve**

Soit le graphe biparti (X, Y) et M^* un couplage maximum. Soit U les nœuds qui ne sont pas incidents à M^* dans X . Soit Z l'ensemble des nœuds atteints par tous les chemins M^* -alternés partant de U . Soient :

- $S = Z \cap X$
- $T = Z \cap Y$

Comme dans la démonstration du théorème du mariage, on montre que $\mathcal{V}(S) = T$.



Soit $K = T \cup (X \setminus S)$, K est une couverture de sommets : toutes les arêtes ont une extrémité dans K car si ce n'est pas le cas, on a une arête du graphe liant $(Y \setminus T)$ à S , or $\mathcal{V}(S) = T$ (contradictoire). On va montrer que K est une couverture minimale et qu'elle vérifie $|K^*| = |M^*|$.

$$|K| = |X \setminus S| + |T| \quad (X \text{ et } Y \text{ étant disjoints}).$$

Supposons qu'il existe $x \in X \setminus S$ tel que x est adjacent par une arête de M^* à un nœud de T , dans ce cas $x \in Z$ et donc $x \in S$ (contradictoire). Donc les arêtes de M^* ont une extrémité dans T ou dans $X \setminus S$ mais pas dans les deux, donc $|K| = |M^*|$, on a donc trouvé K^* est une couverture minimum telle que $|K^*| = |M^*|$.

Notons qu'on n'a pas seulement montré qu'il existe une couverture minimum K^* telle que $|K^*| = |M^*|$, on a aussi vu comment on pouvait la construire connaissant M^* .

6.3.4 Algorithmes pour trouver un couplage maximum et algorithmes pour trouver une couverture minimum

Dans un graphe biparti, en appliquant le théorème de Kőnig et en utilisant les algorithmes produisant un couplage maximum (c.f. ci-après), il est possible de trouver la taille d'une couverture minimum. En revanche pour un graphe quelconque, il existe des algorithmes efficaces pour trouver un couplage maximum, mais on ne connaît pas d'algorithme efficace pour trouver la taille d'une couverture minimum.

6.4 Problème d'assignation des tâches

6.4.1 Version non-pondérée

On suppose avoir des individus compétents pour un certain nombre tâches, comment assigner les tâches à chaque individu pour pouvoir remplir le maximum de tâches possibles? Il s'agit de trouver un couplage maximum dans un graphe biparti (comme le site de rencontre). Il existe des algorithmes efficaces pour réaliser cette tâche.

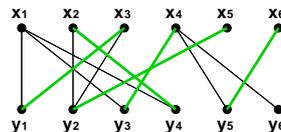
Algorithme hongrois : (version non-pondérée)

- Construire un couplage M du graphe biparti tel qu'on ne puisse plus ajouter d'arêtes sans que deux arêtes soient incidentes au même nœud
- soit $U = \{u \in X \text{ et } u \text{ non-incident à } M^*\}$, si $U = \emptyset$, le couplage est maximum (arrêt)
- $\forall u \in U$, on construit l'ensemble des chemins M -alternés partant de u (étape non-triviale)
 - soit aucun de ces chemins n'est M -augmenté, par théorème le couplage est maximum (arrêt)
 - soit il existe un chemin M -augmenté $u y_a x_b \dots x_l y_m$, et on modifie M selon :

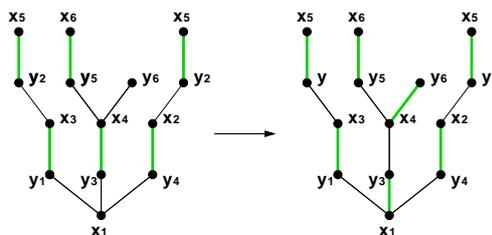
$$M \leftarrow M \setminus \{(y_a, x_b), \dots, (y_k, x_l)\} \cup \{(u, y_a), (x_b, y_c), \dots, (x_l, y_m)\}$$
- on itère le procédé jusqu'à avoir un couplage maximum

exemple:

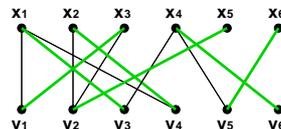
– couplage initial :



– recherche des chemins alternés :



– couplage suivant (et maximum) :



6.4.2 Remarque sur le cas pondéré

Si maintenant on affecte un temps d'exécution d'une certaine tâche par un individu, ou une évaluation du niveau de séduction pour le site de rencontre, on peut le traduire graphiquement en attribuant des pondérations aux liens du graphe bipartis.

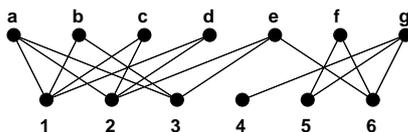
On peut alors chercher un couplage qui réalisera la plus faible durée totale d'exécution des tâches ou la somme des évaluations la plus élevée. Il existe également des algorithmes pour trouver un couplage de poids minimum (ou maximum), dont la méthode hongroise pour les graphes pondérés.

6.5 Parenthèse sur les systèmes de recommandation

La recommandation est une méthode de filtrage visant à ne sélectionner que l'information la plus pertinente pour l'utilisateur. Il peut s'agir de proposer à un client des produits susceptibles de l'intéresser (Amazon), des produits culturels correspondant à ses goûts (Netflix, Last.fm), des individus à qui se connecter sur un réseau social (LinkedIn, FB) etc. Ce domaine connaît une forte expansion avec le développement de services en ligne.

À chaque utilisateur est associé un profil, qui peut être renseigné par l'utilisateur lui-même, ou construit selon ses choix et évaluations passées. Si l'on sait décrire le contenu du produit, on peut alors évaluer l'adéquation avec le profil de l'utilisateur. Comme il est difficile de faire une description exhaustive du contenu des produits, les systèmes s'appuient souvent sur les choix et évaluations des autres utilisateurs pour recommander des produits choisis par des clients ayant des profils assez similaires. C'est le principe du **filtrage collaboratif**. Dans ce contexte, le problème peut être représenté à l'aide d'un graphe biparti (clients, produits), éventuellement pondéré (évaluation), et il s'agit alors de prévoir les liens pertinents ou leur poids.

exemple: Prenons un exemple simple sans évaluation :



Étant donné la structure du réseau biparti, on recommanderait à **d** le produit **3** mais pas le **4**.

En pratique ce type de problème est souvent résolu à l'aide de techniques d'**apprentissage** (*machine learning*). L'idée essentielle est de construire un modèle pour déterminer à partir de ce que l'on sait d'un utilisateur les produits qui peuvent l'intéresser ; dans ce modèle les informations (achats antérieurs, profil...) sont pondérées, et les poids peuvent être appris à l'aide des informations sur les autres utilisateurs : si par exemple tous les utilisateurs ayant acheté **2** ont acheté **3**, alors le poids de l'information "**d** a acheté **2**" devra être très élevé pour évaluer l'intérêt que **d** a pour **3**. De nombreuses techniques existent pour réaliser les tâches de ce type (régression logistique, *Support Vector Machine*, réseaux neuronaux...).

Chapitre 7

Coloriages d'arêtes

Dans tout ce chapitre, on supposera des graphes **sans boucle**.

7.1 Définitions de base

* *Définition*

Un **coloriage des arêtes** d'un graphe en k couleurs est l'assignation à chaque arête d'une couleur $1, 2, \dots, k$.

* *Définition*

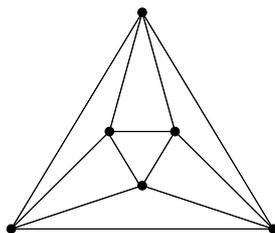
Un coloriage est **propre** si deux arêtes adjacentes sont toujours de couleurs différentes.

Remarque : trouver un coloriage propre en k couleurs consiste à séparer les arêtes en k couplages distincts.

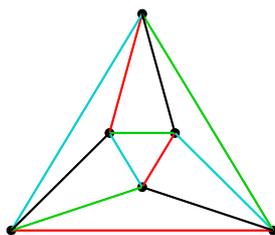
* *Définition*

L'indice chromatique d'un graphe G noté $\chi'(G)$ est le nombre minimal de couleurs nécessaire pour obtenir un coloriage propre des arêtes de G .

exemple: Le graphe suivant a un indice chromatique de 4.



En effet :



Application : Organisation de tournois toutes rondes (*round-robin tournaments*). Tous les participants d'un tournoi doivent affronter tous les autres participants, éventuellement plusieurs fois (et pas nécessairement le même nombre de fois). La problème revient à trouver un coloriage propre minimum au graphe dont les sommets sont les participants au tournoi et les arêtes les confrontations entre ces participants. Toutes les confrontations de la même couleur se déroulent au même moment.

7.2 Indice chromatique et degré maximum

Remarque : Soit δ_{max} le degré maximum, alors $\chi' \geq \delta_{max}$.

• Théorème

Théorème de König : Pour un (multi-)graphe biparti, $\chi' = \delta_{max}$.

◦ Preuve

Soit un graphe biparti G , de bipartition (X, Y) et de degré maximum δ_{max} .

- On augmente G pour en faire un graphe δ_{max} -régulier : supposons $|X| \leq |Y|$, on ajoute des nœuds à X jusqu'à égaliser. Tant qu'il existe un nœud $x \in X$ de degré $< \delta_{max}$, il existe $y \in Y$ de degré $< \delta_{max}$ (sinon il n'y aurait pas le même nombre d'arêtes incidentes aux deux ensembles), on ajoute une arête entre x et y . On itère ce procédé jusqu'à obtenir un graphe δ_{max} -régulier.
- On applique alors le corollaire du théorème du mariage à ce graphe biparti G' δ_{max} -régulier : il existe au moins un couplage parfait pour ce graphe, on choisit un de ces couplages. On colore alors les arêtes de ce couplage dans la couleur c_1 , puis on retire ces arêtes de G' , le graphe obtenu est $(\delta_{max} - 1)$ -régulier, on itère alors la procédure en colorant à chaque itération les arêtes dans une nouvelle couleur c_i .
- Au terme de cette procédure on a un coloriage propre à δ_{max} couleurs de G' , ce coloriage est aussi un coloriage propre pour les arêtes de G , donc $\chi' = \delta_{max}$.

• Théorème

Théorème de Vizing (1964) : Pour un graphe simple, $\chi' = \delta_{max}$ ou $\chi' = \delta_{max} + 1$.

◦ Preuve

Soit un graphe simple de degré maximum δ_{max} , il suffit de montrer que $\chi' \leq \delta_{max} + 1$, on procède par récurrence sur le nombre d'arêtes du graphe.

Pour un graphe à δ_{max} arêtes, la propriété est vérifiée. On suppose pour tout graphe comptant à moins de m arêtes et de degré maximum δ_{max} , il existe un coloriage propre à au plus $\delta_{max} + 1$ couleurs. Soit G à m arêtes, on lui retire une arête xy_0 (qui ne modifie pas son degré maximum), le graphe obtenu admet donc un coloriage propre à au plus $\delta_{max} + 1$ couleurs.

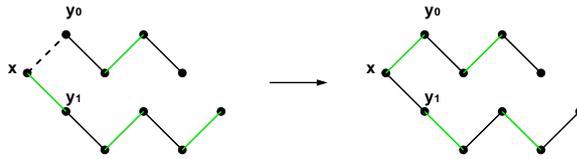
Si ce coloriage propre a exactement δ_{max} couleurs, on ajoute xy_0 dans une nouvelle couleur et le problème est résolu. Considérons le cas où le graphe admet un coloriage propre à $\delta_{max} + 1$ couleurs exactement, comme le coloriage a plus de couleurs que δ_{max} , chaque nœud a nécessairement au moins une couleur du coloriage

manquante. Il manque au nœud x au moins une couleur c .

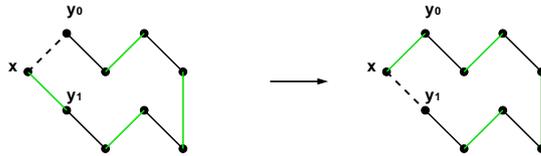
Pour continuer la preuve on introduit la notion de **chaîne de Kempe** : une chaîne de Kempe de couleur a et b , notée $H(a, b)$, est une composante connexe du graphe formée par les arêtes de couleurs a et b . Comme on l'a vu dans le chapitre 6, une telle chaîne ne peut être qu'un cycle de longueur paire ou un chemin alterné ouvert terminant par des nœuds de degré 1.

Plusieurs cas sont possibles :

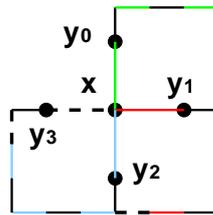
- La couleur absente c pour x est manquante pour y_0 , dans ce cas on ajoute cette arête colorée en c et le problème est résolu.
- Si la couleur c_0 manquante pour y_0 n'est pas c , alors il existe y_1 tel que xy_1 soit de couleur c_0 . On examine alors la chaîne de Kempe $H(c, c_0)$ commençant par l'arête xy_1 :
 - Soit cette chaîne est disjointe de la chaîne terminant par l'arête de couleur c en y_0 . Dans ce cas on recolore les arêtes c_0 de cette chaîne en c et les arêtes c en c_0 , alors c_0 devient manquante en x , on peut alors colorer xy_0 en c_0 et le problème est résolu.



- Soit cette chaîne termine par l'arête de couleur c en y_0 . On ajoute alors l'arête xy_0 en c_0 et on décolore l'arête xy_1 , et on fait le même raisonnement : si c_1 la couleur manquante de y_1 , soit $c_1 = c$ et dans ce cas le problème est résolu, soit $c_1 \neq c$ et on examine les chaînes de Kempe $H(c, c_1)$.

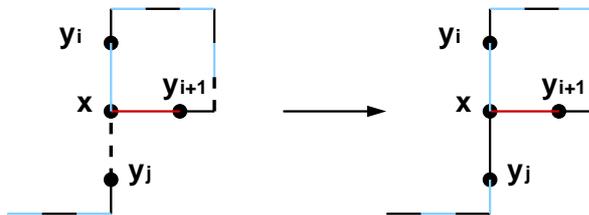


- On itère ce raisonnement :



- Si nous atteignons le dernier voisin non-examiné y_k , dans ce cas sa couleur manquante ne peut être qu'une couleur manquante de x sinon elle aurait été vue au cours du processus. Si on n'atteint pas ce dernier voisin, c'est donc qu'il exist une valeur j associée à un voisin y_j telle que la couleur manquante c_j qui lui est associée est identique à une couleur c_i déjà rencontrée. Considérons la chaîne de Kempe $H(c, c_i)$ partant de xy_i jusqu'à y_{i+1} : y_j n'appartient pas à cette chaîne car tous le nœuds de la chaîne ont la couleur c_i excepté y_{i+1} dont l'arête xy_{i+1} qui était c_i vient d'être décolorée (et $j \neq i + 1$ pour cette raison). Mais on peut construire une chaîne $H(c, c_i)$ partant de y_j car ce nœud a la couleur c , et cette seconde chaîne est nécessairement disjointe, on permute donc

les colorations de cette chaîne et on peut colorer xy_j en c et résoudre le problème.



• **Remarque algorithmique :**

Il existe un algorithme efficace pour calculer χ' .

\iff

Il existe un algorithme efficace pour décider s'il existe une couverture de sommets de taille $\leq k$.

\iff

Il existe un algorithme efficace pour décider si un graphe est hamiltonien.

Il n'y a pas d'algorithme connu pour décider si, pour un graphe simple donné $\chi' = \delta_{max}$.

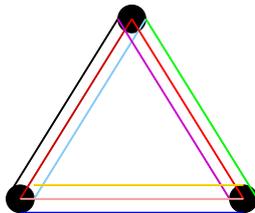
• **Théorème**

Théorème de Vizing généralisé au cas des multigraphes :

Pour un multigraphe, s'il y a au plus m arêtes entre deux nœuds :

$$\delta_{max} \leq \chi' \leq \delta_{max} + m$$

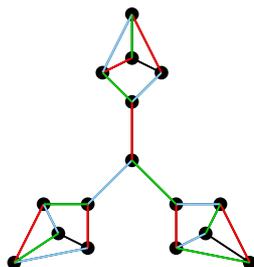
exemple: pour ce graphe, $\chi' = \delta_{max} + m$:



7.3 Relation avec les couplages

Si un couplage maximum est “petit”, il faudra “beaucoup” de couplages maximum distincts pour arriver à colorier l’ensemble des arêtes du graphe. On peut utiliser cette remarque qualitative pour montrer que $\chi' = \delta_{max} + 1$ dans des cas particuliers.

exemple: Dans ce graphe un couplage maximum a 7 arêtes, comme le graphe compte 24 arêtes, il n’est pas possible de trouver un coloriage propre à $\delta_{max} = 3$ couleurs, donc $\chi' = 4$.



Remarque : attention, cela ne fournit pas pour autant une condition nécessaire et suffisante, car s’il on supprime le sommet en haut du graphe précédent, nous obtenons un graphe à 21 arêtes dont le coloriage propre à 3 couleurs est impossible, alors qu’un couplage maximum a toujours 7 arêtes.

7.4 Problème de la confection d’horaire

On peut représenter ce problème à l’aide d’un graphe biparti pondéré (enseignants / classes). On suppose qu’un enseignant peut enseigner au maximum à une classe par heure, et qu’une classe ne peut avoir qu’un enseignant dans l’heure. Un emploi du temps pour une heure revient à chercher un couplage enseignant-classe.

Notre problème revient donc à trouver le moins de couplages possibles (c’est-à-dire un coloriage propre de taille minimum) pour associer enseignants et classes. D’après le théorème de Kőnig, il existe donc un coloriage propre de taille δ_{max} , donc si une classe n’a pas cours pendant plus de k heures et qu’un professeur n’enseigne pas plus de k heures, on doit pouvoir trouver un horaire sur k heures qui convienne.

Il existe un algorithme efficace pour résoudre ce problème.

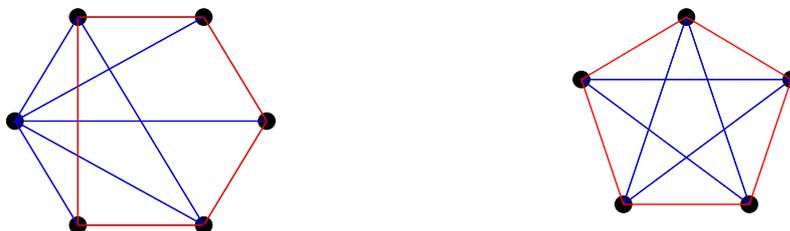
Chapitre 8

Cliques, ensembles indépendants et l'impossible désordre

8.1 Le théorème de l'amitié

Jeu de Sim : Imaginons des points dans un plan, les deux joueurs ont un crayon de couleur différente et tracent chacun tour à tour une arête entre deux points. Le premier à compléter un triangle de sa couleur a perdu.

exemple: Avec 6 points dans le plan, il n'y a jamais de partie nulle. Avec 5 points en revanche, c'est possible.



• Théorème

Théorème de l'amitié : Parmi six personnes on en trouve toujours soit trois qui se connaissent l'une l'autre, soit trois qui sont étrangères l'une à l'autre. Autrement dit, tout coloriage à deux couleurs du graphe complet à six nœuds crée un triangle monochrome.

◦ Preuve

On fixe les couleurs à bleu et rouge. Soit v un des 6 nœuds de degré 5, parmi ses arêtes 3 au moins sont de la même couleur : fixons AV , BV et CV sont bleues.

- Soit AB , BC et CA sont rouges et donc ABC est rouge.
- Soit une arête au moins parmi AB , BC et CA est bleue et donc ABV , BCV ou ACV est bleu.

8.2 Ensembles indépendants et couvertures de sommets et cliques

* Définition

Un **ensemble indépendant** d'un graphe est un ensemble de nœuds deux à deux non-adjacents.

* Définition

Un **ensemble indépendant maximum** d'un graphe est un ensemble indépendant dont le nombre de nœuds est maximum.

• **Théorème**

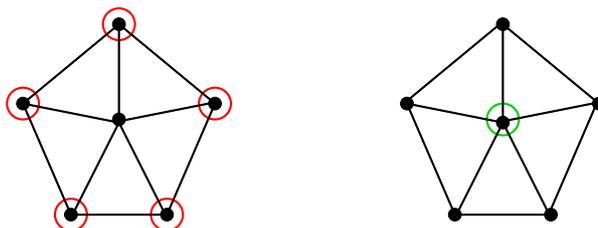
Un ensemble de nœuds est indépendant si et seulement si son complémentaire est une couverture de sommets.

◦ **Preuve**

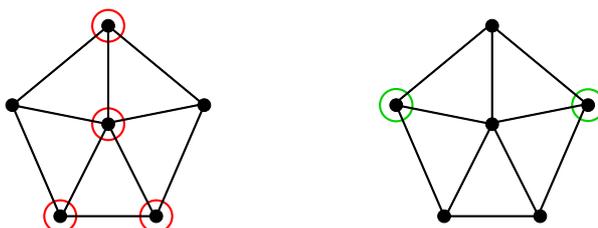
Soit V l'ensemble des sommets d'un graphe et S un ensemble indépendant, alors :

$$\begin{array}{c}
 S \text{ est indépendant} \\
 \Updownarrow \\
 \text{aucune arête n'a ses deux extrémités dans } S \\
 \Updownarrow \\
 \text{chaque arête a au moins une extrémité dans } V \setminus S \\
 \Updownarrow \\
 V \setminus S \text{ est une couverture de sommets}
 \end{array}$$

exemple: À gauche : couverture de sommets, à droite ensemble indépendant associé.



À gauche : couverture de sommets minimum, à droite : ensemble indépendant maximum.



Corollaire : Pour un graphe à n nœuds, si S^* est un ensemble indépendant maximum et K^* une couverture minimum, alors :

$$|S^*| + |K^*| = n$$

Et trouver un ensemble indépendant maximum est aussi difficile que de trouver une couverture minimum. Il n'y a aucun algorithme efficace connue pour chacune de ces deux tâches.

* Définition

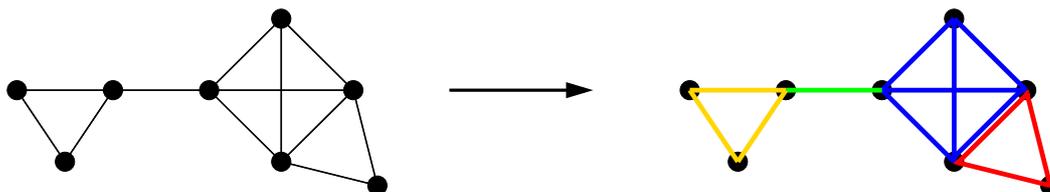
Une **clique** d'un graphe est un ensemble de nœuds deux à deux adjacents, autrement dit, c'est un sous-graphe complet.

* Définition

Une **clique maximum** est une clique dont le nombre de nœuds est maximum.

exemple:

Ce graphe comporte 1 clique maximum de taille 4, 2 de taille 3 et 1 de taille 2.



• **Théorème**

Un ensemble est indépendant dans un graphe simple si et seulement si il est une clique du graphe complémentaire.

◦ **Preuve**

Soit G un graphe et S une clique de G , alors :

$$\begin{array}{c}
 S \text{ est une clique} \\
 \Updownarrow \\
 \forall x, y \in S, \text{ il existe une arête entre } x \text{ et } y \\
 \Updownarrow \\
 \forall x, y \in S, x \text{ n'est pas adjacent à } y \text{ dans le complémentaire de } G \\
 \Updownarrow \\
 S \text{ est un ensemble indépendant dans le complémentaire de } G
 \end{array}$$

Remarques :

- On a utilisé le fait que x et y sont adjacents dans $G \Leftrightarrow x$ et y ne sont pas adjacents dans le complémentaire de G .
- Donc trouver une clique maximum est aussi difficile que de trouver un ensemble indépendant maximum et que de trouver une couverture de sommets minimum.

8.3 Théorème de Ramsey

Un triangle est une clique à 3 nœuds, il existe une généralisation du théorème de l'amitié pour des cliques de taille quelconque et un nombre arbitraire de couleurs : le théorème de Ramsey.

• **Théorème**

Théorème de Ramsey (1930) : Pour tout couple d'entiers p et q , il existe un plus petit entier r tel que tout graphe à r nœuds contient soit une clique à p nœuds, soit un ensemble indépendant à q nœuds. On le note $R(p, q)$, c'est le nombre de Ramsey.

Cela revient à dire que pour $R(p, q)$ est la taille minimum du graphe complet tel que tout coloriage à deux couleurs c_1, c_2 du graphe contient soit une clique de taille p de couleur c_1 , soit une clique de taille q de couleur c_2 .

Remarques :

- le théorème de l'amitié est alors contenu dans l'égalité $R(3, 3) = 6$
- on note que $R(1, q) = R(p, 1) = 1$ et $R(2, q) = q, R(p, 2) = p$
- la formulation avec les couleurs met en évidence que $R(a, b) = R(b, a)$

• **Théorème**

Théorème d'Erdős-Szekeres (1935) : pour $p, q \geq 2$, on a

$$R(p, q) \leq R(p - 1, q) + R(p, q - 1)$$

◦ **Preuve**

On considère le graphe complet à $n = R(p - 1, q) + R(p, q - 1)$ sommets, qu'on appelle G et dont on a un coloriage des arêtes en rouge et bleu. Nous devons donc montrer que $n \geq R(p, q)$, c'est-à-dire que G contient soit une clique rouge de taille p , soit une clique bleue de taille q (ou l'inverse).

Soit un nœud v quelconque. On note alors :

- A l'ensemble des voisins de v par une arête rouge
- B l'ensemble des voisins de v par une arête bleue

Le graphe étant complet, v a $n - 1 = R(p - 1, q) + R(p, q - 1) - 1$ voisins, donc

$$|A| + |B| + 1 = R(p - 1, q) + R(p, q - 1)$$

Deux cas sont possibles :

- a) Soit $|A| \geq R(p - 1, q)$ (v a au moins $R(p - 1, q)$ voisins par des arêtes rouges).
Comme v est lié aux éléments de A par des arêtes rouges, alors en ajoutant v à A , on a un sous-graphe de G qui contient soit une clique rouge de taille p , soit une clique bleue de taille q .
- b) Soit $|B| \geq R(p, q - 1)$ (v a au moins $R(p, q - 1)$ voisins par des arêtes bleues).
Comme v est lié aux éléments de B par des arêtes bleues, alors en ajoutant v à B , il existe un sous-graphe de G qui contient soit une clique rouge de taille p , soit une clique bleue de taille q .

Remarque :

Avec l'inégalité d'Erdős-Szekeres, on a montré implicitement que si $R(p, q - 1)$ et $R(p - 1, q)$ existent, alors $R(p, q)$ existe. Alors sachant que $R(p, 1) = R(1, q) = 1$, par récurrence tout $R(p, q)$ existe, par exemple :

- $R(2, 2) \leq R(1, 2) + R(2, 1) = 2$
- $R(3, 2) \leq R(2, 2) + R(3, 1) \leq 3$
- $R(3, 3) \leq R(3, 2) + R(2, 3) \leq 6$

Sur les combinatoires :

- $\binom{n}{k}$ désigne le nombre de combinaisons à k éléments dans un ensemble qui en comprend n , sans ordre, sans remise et sans répétition. Par exemple le nombre de mains de 5 cartes possibles pour un jeu de 52 cartes est $\binom{52}{5}$.
- Pour calculer ce type de quantité, on utilise : $\binom{n}{k} = \frac{n!}{(n-k)!k!}$.
- On dispose également de la formule du triangle de Pascal :

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

- **Corollaire de l'inégalité d'Erdős-Szekeres :** $\forall p, q \geq 2$

$$R(p, q) \leq \binom{p+q-2}{p-1}$$

◦ **Preuve**

Par récurrence sur $p+q$. On peut vérifier que l'inégalité est vérifiée pour $p+q \leq 5$. Supposons que l'inégalité soit vérifiée $\forall p, q$ tels que $5 \leq p+q < r+s$. Alors pour $p+q = r+s$, on applique le théorème d'Erdős-Szekeres :

$$R(p, q) \leq R(p, q-1) + R(p-1, q)$$

$$\Rightarrow R(p, q) \leq \binom{p+(q-1)-2}{p-1} + \binom{(p-1)+q-2}{(p-1)-1} = \binom{p+q-3}{p-1} + \binom{p+q-3}{p-2}$$

On applique alors la formule du triangle de Pascal :

$$R(p, q) \leq \binom{p+q-2}{p-1}$$

On connaît très peu de nombres de Ramsey, parmi ceux-là :

p, q	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	2	3	4	5	6	7
3	1	3	6	9	14	18	23
4	1	4	9	18			
5	1	5	14				
6	1	6	18				
7	1	7	23				

8.4 Cliques et densité

• Théorème

Théorème de Turán (1941) : Si G est un graphe simple à n nœuds ne contenant aucune clique à $r + 1$ nœuds, alors G compte au plus $(1 - \frac{1}{r})\frac{n^2}{2}$ arêtes.

Remarques :

- pour $r = 2$ et $n = 6$, on retrouve le théorème de l'amitié
- pour un graphe complet ($r = n - 1$), on peut vérifier que $(1 - \frac{1}{r})\frac{n^2}{2} < \frac{n(n-1)}{2}$

8.5 Théorie de Ramsey

Le théorème de Ramsey admet une formulation plus générale avec un nombre de couleurs quelconques.

• Théorème

Théorème de Ramsey : Pour tout p -uplet d'entiers q_1, \dots, q_p , il existe un plus petit entier noté $R(q_1, \dots, q_p)$ tel que tout coloriage à p couleurs du graphe complet de taille $R(q_1, \dots, q_p)$ contient soit une clique de taille q_1 de couleur c_1 , soit une clique de taille q_2 de couleur c_2 , ..., soit une clique de taille q_p de couleur c_p .

Schématiquement, le théorème de Ramsey traduit l'idée qu'à partir d'une certaine taille il apparaît une forme d'ordre dans un ensemble d'objets. Le désordre complet à n'importe quelle échelle n'est pas possible. La théorie de Ramsey cherche alors à répondre à cette question générale : "combien d'éléments sont nécessaires à une certaine structure pour que telle propriété soit vérifiée ?" Plusieurs théorèmes admettent des formulations de ce type, on en cite ci-dessous quelques exemples.

• Théorème

Théorème de Schur (1916) : Pour tout entier k , il existe un entier r_k tel que toute partition des nombres $1, 2, \dots, r_k$ en k classes, une de ces classes comprend x, y et z (pas nécessairement distincts) tels que $x + y = z$.

exemple : $r_3 = 14$

• Théorème

Théorème de Van der Waerden (1927) : Pour tout couple d'entiers p et q , il existe un entier minimum $W(p, q)$ tel que toute partition des nombres $1, 2, \dots, W(p, q)$ en p classes comprend une classe dans laquelle il y ait au moins q entiers en progression arithmétique, c'est-à-dire de la forme $a + r = b, b + r = c, \dots$

exemple : $W(2, 3) = 9$

Chapitre 9

Coloriage de sommets

9.1 Définitions de base

Risque Seveso :

Une usine qui doit stocker un grand nombre de produits chimiques potentiellement dangereux doit respecter des normes de sécurité, en particulier elle doit stocker en des lieux séparés les produits qui mélangés pourraient causer une explosion. Comment faut-il stocker les produits ?

En termes de graphes, on représente un produit chimique par un sommet du graphe, une association potentiellement dangereuse entre 2 produits par une arête, et une couleur de sommet représente l'entrepôt dans lequel le produit est stocké. Il faut alors ne pas avoir deux nœuds de la même couleur adjacents, combien de couleurs sont nécessaires ?

* Définition

Un **coloriage de nœuds** est l'affectation à chaque nœud d'une couleur. Le coloriage est **propre** si des nœuds adjacents sont de couleurs différentes.

* Définition

Un coloriage **minimum** est un coloriage propre qui emploie un nombre minimal de couleurs.

* Définition

Le **nombre chromatique** d'un graphe G , noté $\chi(G)$, est le nombre de couleurs d'un coloriage minimum.

• Théorème

$$|\text{clique maximum}| \leq \chi \leq \delta_{max} + 1$$

◦ Preuve

On démontre chaque inégalité :

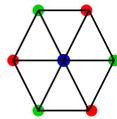
- $|\text{clique maximum}| \leq \chi$: pour une clique à k nœuds, il faut nécessairement que les k sommets soient de couleurs différentes.
- $\chi \leq \delta_{max} + 1$: par récurrence sur le nombre de nœuds, on raisonne à δ_{max} fixé. Pour les graphes de degré maximum δ_{max} contenant au plus $\delta_{max} + 1$ nœuds, la propriété est vérifiée, supposons que ce

soit vrai pour un graphe à $n - 1$ nœuds. On raisonne à δ_{max} fixé dans un graphe à n nœuds, en retirant un nœud u quelconque, on a un graphe à $n - 1$ nœuds de degré $\leq \delta_{max}$, qu'on colorie avec au plus $\delta_{max} + 1$ couleurs. En ajoutant u qui a au plus δ_{max} voisins, il reste au moins une couleur disponible pour colorier u .

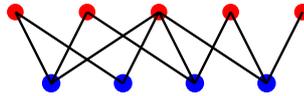
Malgré ces bornes, on ne connaît pas d'algorithme efficace pour calculer systématiquement χ .

exemples :

- $\chi = 3, \delta_{max} = 6$



- $\chi = 2$ (graphe biparti)



- $\chi = \delta_{max} + 1 = 4$ (clique)



9.2 Polynôme chromatique

Soit $\pi_k(G)$ le nombre de coloriage propres à k couleurs du graphe G .

Rappel : $G \cdot e$ représente le graphe G dont on fusionne les deux extrêmités de l'arête e .

• Théorème

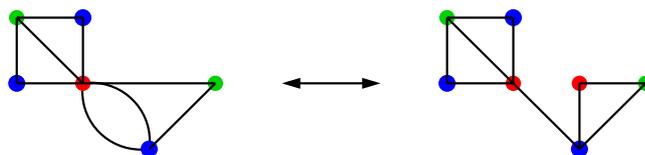
Pour toute arête e du graphe :

$$\pi_k(G) = \pi_k(G - e) - \pi_k(G \cdot e)$$

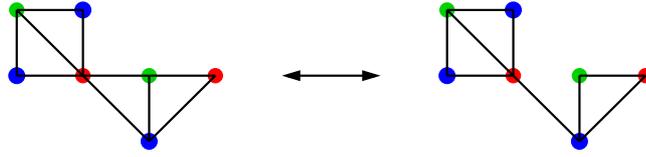
◦ Preuve

Soient u et v les extrêmités de e . Dans $G - e$, on peut distinguer deux types de coloriage propres :

- Les coloriage propres où u et v sont de la même couleur, on peut construire une bijection entre l'ensemble de ces coloriage propres et ceux de $G \cdot e$, comme sur le schéma :



- Les coloriage propres où u et v sont de couleurs différentes, on peut construire une bijection triviale entre l'ensemble de ces coloriage propres et ceux de $G - e$:



$$\implies \pi_k(G - e) = \pi_k(G \cdot e) + \pi_k(G)$$

• **Théorème**

Corollaire (Birkhoff, 1912) : Pour un graphe G à n nœuds, $\pi_k(G)$ est un polynôme unitaire en k , de degré n , de terme constant nul et dont les coefficients sont entiers et alternent de signes.

Exemples :

- Pour n nœuds isolés et k couleurs, on peut attribuer n'importe quelle couleur à n'importe quel nœud, donc $\pi_k = k^n$ coloriage propre.
- Pour un graphe complet à n nœuds et k couleurs, on doit attribuer des couleurs différentes à chacun des n nœuds, donc $\pi_k = k(k - 1)(k - 2)\dots(k - n + 1)$

◦ **Preuve**

Par récurrence sur le nombre d'arêtes m . Pour $m = 0$, on a vu ci-dessus que la propriété était vérifiée, on suppose maintenant qu'elle est vérifiée pour tout graphe G' avec un nombre d'arêtes $< m$, autrement dit on peut mettre $\pi_k(G')$ sous la forme :

$$\pi_k(G') = k^n + \sum_{i=1}^{n-1} (-1)^{n-i} a_i k^i \text{ avec } a_i \in \mathbb{N}$$

Pour G à m arêtes :

$$\pi_k(G) = \pi_k(G - e) - \pi_k(G \cdot e)$$

Or $G - e$ et $G \cdot e$ ont $m - 1$ arêtes (les éventuelles arêtes multiples de $G \cdot e$ peuvent être ramenées à une arête simple sans changer le problème), $G - e$ a n sommets et $G \cdot e$ en a $n - 1$, donc

$$\begin{aligned} \pi_k(G) &= k^n + \sum_{i=1}^{n-1} (-1)^{n-i} a_i k^i - \left(k^{n-1} + \sum_{i=1}^{n-2} (-1)^{n-1-i} b_i k^i \right) \text{ avec } a_i, b_i \in \mathbb{N} \\ \implies \pi_k(G) &= k^n - (a_{n-1} + 1)k^{n-1} + \sum_{i=1}^{n-2} (-1)^{n-i} (a_i + b_i) k^i \end{aligned}$$

On appelle ce polynôme le **polynôme chromatique** de G , et il n'y a pas d'algorithme efficace pour le calculer.

Chapitre 10

Graphes planaires

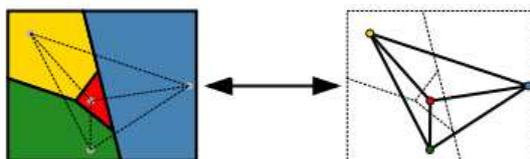
10.1 Question du coloriage d'une carte

Considérons une carte dont on souhaite colorier les pays de manière à ce que deux pays adjacents ne soient pas de la même couleur. Quel est le nombre de couleurs minimum pour réaliser ce coloriage ?

Cette question a été posée dès 1852 par Francis Guthrie pour le coloriage des comtés (*counties*) d'Angleterre, observant que 4 couleurs sont suffisantes dans ce cas précis.



On considère le cas où chaque territoire est connexe (d'un seul tenant), et deux territoires sont colorés différemment s'ils partagent une frontière commune (pas seulement un point). Alors la question peut-être comprise comme un problème d'évaluation d'un nombre chromatique : un territoire est un sommet du graphe, une arête signifie que deux territoires partagent une frontière. Mais les graphes considérés ont alors une propriété très particulière : il est possible de les représenter dans le plan de manière à ce que les arêtes ne s'intersectent qu'aux extrémités...



10.2 Notion de graphe planaire

* Définition

Un graphe est **planaire** s'il possède une représentation dans le plan dont les arêtes ne se touchent pas, sauf à leurs extrémités.

* Définition

Une **subdivision** d'un graphe est le remplacement d'une arête par un chemin.

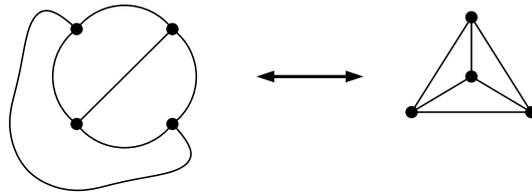
Remarques :

- un sous-graphe d'un graphe planaire est planaire
- une subdivision d'un graphe non-planaire est non-planaire

• **Théorème**

Théorème de Fáry (1948) : Tout graphe planaire simple peut être représenté en n'utilisant que des arêtes droites.

exemple:



10.3 Graphes non-planaire

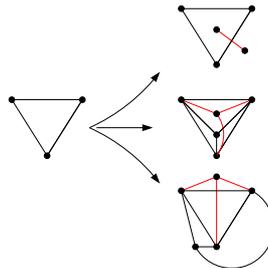
• **Théorème**

Le graphe complet à 5 nœuds (K_5) est non-planaire.

◦ **Preuve**

Supposons qu'il existe une représentation planaire avec les sommets v_1, v_2, v_3, v_4, v_5 de K_5 , on examine le triangle v_1, v_2, v_3 .

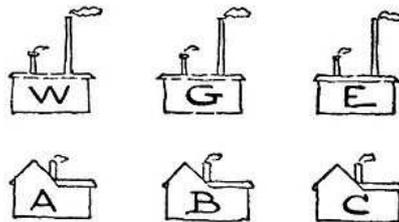
- Si v_4 est intérieur et v_5 extérieur au cycle, alors (v_4, v_5) intersecte le triangle.
- Si v_4 et v_5 sont intérieurs, alors v_4 doit être intérieur à un des trois triangles v_1, v_2, v_5 ; v_1, v_3, v_5 ou v_3, v_2, v_5 . Dans ce cas le dernier sommet v_i est lui extérieur à ce triangle et (v_4, v_i) l'intersecte.
- Si v_4 et v_5 sont extérieurs, alors on doit avoir un triangle v_1, v_2, v_4 ; v_1, v_3, v_4 ou v_3, v_2, v_4 tel que v_5 est extérieur à ce triangle et le dernier nœud v_i est intérieur. Donc (v_5, v_i) intersecte ce triangle.



• **Théorème**

Le graphe biparti complet à 3+3 nœuds ($K_{3,3}$) est non-planaire.

Remarque : Ce théorème répond à “l’énigme des 3 maisons” d’Henry Dudeney, dans laquelle on veut relier trois maisons à l’eau, le gaz et l’électricité, sans que les alimentations se croisent (c.f. schéma ci-dessous). Ce n’est donc pas possible dans le plan.

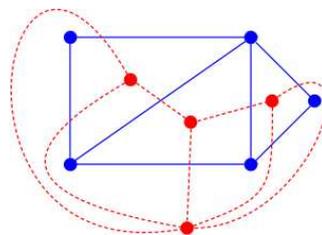


• **Théorème**

Théorème de Kuratowski (1930) : Un graphe est non-planaire si et seulement si il contient K_5 , $K_{3,3}$ ou une subdivision de ceux-ci.

Complément :

Une autre définition utile pour raisonner sur les graphes planaires est la notion de **graphe dual**. Le graphe planaire permet de définir des **faces**, c’est-à-dire des surfaces du plan délimitées par les arêtes et les sommets du graphe. On définit le dual comme le graphe tel qu’un sommet représente une face et une arête l’existence d’une frontière commune entre deux faces. Ce graphe est également planaire.



10.4 Retour au coloriage d’une carte

Kempe en 1879 pensait avoir démontré que 4 couleurs suffisaient, mais sa démonstration sera prouvée incorrecte en 1891, en revanche le plan de la démonstration sera conservé. On ne savait alors montrer que le théorème des 5 couleurs, l’idée que 4 couleurs sont suffisantes conservait le statut de conjecture.

Ce n’est qu’en 1976 qu’Appel et Haken démontrent le théorème des 4 couleurs. Leur démonstration utilise un dénombrement de configurations problématiques, en nombre trop important pour être réalisé manuellement, il s’agissait alors d’une des premières preuves assistées par ordinateur.