

# Rapport de Projet de Calcul Quantique : PageRank

Théodore CHAPUIS-CHKAIBAN

June 2020

## 1 Introduction et objectifs

Le calcul Quantique est un domaine passionnant à la croisée des mathématiques, de la physique et de l'informatique. Ces vingt dernières années, l'essor de l'informatique quantique a été considérable et s'est soldé en plusieurs prouesses scientifiques. Google a récemment affirmé avoir atteint la suprématie quantique en stabilisant un ordinateur de plus de 50 Qbits : cette prouesse technique est l'aboutissement de plusieurs dizaines d'années de recherches dans le domaine de la physique théorique mais aussi une compréhension aboutie et mature de technologies qui, jusqu'à très récemment, se retrouvaient reléguées au stade de lubies futuristes, sans beaucoup d'avenir en somme.

Aujourd'hui, notre maîtrise technique des ordinateurs quantiques est telle qu'elle a supplanté notre progression dans le développement d'algorithmes quantiques : l'enjeu des prochaines années sera d'emboîter effectivement le pas et de produire un système cohérent et efficace d'algorithmes de calcul quantique.

L'idée de ce compte rendu est ainsi de proposer une nouvelle méthode de calcul du PageRank de Google, méthode qui serait plus efficace en complexité ( $O(\log(n))$ ) contre  $O(\sqrt{n})$  jusqu'alors).

Nous ne ferons qu'esquisser ici l'idée de cette méthode à laquelle notre travail de cette année a pu aboutir. Le projet se poursuivant en 2ème année, nous avons l'espoir que cette méthode puisse nous permettre d'obtenir des résultats dignes d'intérêt à terme.

## 2 Compte rendu du travail de l'année

Nous rendons-compte ici des étapes principales de notre travail sur l'année 2020 :

Tâche accomplie	Début	Fin
Mise à niveau en Calcul Quantique	Octobre 2019	Décembre 2019
Définition d'une problématique et d'un sujet de Recherche	Décembre 2019	Février 2020
Développement d'algorithmes Quantiques et Classique de PageRank	Décembre 2019	Février 2020
Elargissement de notre sujet de recherche (Graph Fourier Transform)	Février 2020	Avril 2020
Adaptation de la transformée de Fourier au problème considéré	Avril 2020	Mai 2020

FIGURE 1 – Grandes étapes du projet de Calcul Quantique (PageRank) de l'année.

## Table des matières

<b>1</b>	<b>Introduction et objectifs</b>	<b>1</b>
<b>2</b>	<b>Compte rendu du travail de l'année</b>	<b>1</b>
<b>3</b>	<b>Introduction au PageRank</b>	<b>3</b>
3.1	PageRank Classique . . . . .	3
3.2	PageRank Quantique . . . . .	4
3.3	Points de similitudes entre versions classiques et quantiques. . . . .	4
<b>4</b>	<b>Apport de la théorie spectrale des Graphes</b>	<b>5</b>
4.1	Contextualisation . . . . .	5
4.2	Introduction à la théorie . . . . .	5
4.3	Transformée de Fourier pour un graphe non orienté . . . . .	7
4.4	Généralisation aux graphes orientés . . . . .	8
<b>5</b>	<b>Mise au point d'un algorithme Quantique de résolution du PageRank.</b>	<b>10</b>
5.1	Présentation Sommaire . . . . .	10
<b>6</b>	<b>Conclusion et remerciements</b>	<b>10</b>
<b>7</b>	<b>Bibliographie</b>	<b>11</b>

## 3 Introduction au PageRank

### 3.1 PageRank Classique

Le PageRank est un système introduit par Google qui est à la base du moteur de recherche internet. L'intérêt du PageRank est d'attribuer un score aux pages du Web pour les classer par ordre de pertinence lors d'une recherche internet.

La méthode classique du PageRank définit le concept d'importance formellement de la manière suivante :

**Définition 1** *L'importance d'un noeud d'un graphe orienté  $G=(V,E)$  est définie par l'équation :*

$$I(P_i) = \sum_{j \in B_i} \frac{I(P_j)}{\deg(P_j)}$$

*Avec  $P_i$  la page  $i$ ,  $\deg(P_i)$  le nombre de liens sortant de la page  $i$  et  $B_i$  l'ensemble des pages pointant sur la page  $i$ .*

Cette équation est cohérente avec le concept d'importance de page Web : plus une page est référencée par des pages importantes, plus celle-ci a, elle-même, de l'importance. Cette importance est toutefois pondérée par le nombre de liens externes de chaque page qui fait référence à la page  $i$  (si on fait peu de références autres qu'à la page  $i$ , tout en ayant une importance élevée, on confère beaucoup d'importance à la page  $i$ ).

Cette équation est récursive ( $I$  est définie en fonction des autres valeurs de l'importance), on voit donc qu'avec cette équation seulement, l'importance n'est pas définie de manière unique.

Soit alors  $I$  le vecteur de  $\mathbb{R}^N$  (avec  $N$  le nombre de page du web) et  $H$  une matrice telle que  $H_{i,j} = \frac{1}{\deg(P_j)} \delta_{j \in B_i}$ , il vient :

$$HI = I$$

Qui est une équation aux valeurs propres sur le nombre de pages du web. Comme la matrice  $H$  est typiquement de taille  $10^9 * 10^9$ , il est impensable de diagonaliser directement  $H$  (en tout cas par des méthodes classiques). C'est ainsi que, bien souvent, on travaille sur une forme légèrement différente de  $H$  (matrice primitive) qui puisse nous permettre d'appliquer le théorème de Perron Frobenius pour approximer le vecteur propre associé à la valeur propre 1 par exponentiation successives.

Pour rendre  $H$  primitive, il faut la rendre à coefficients positifs et irréductible (il n'existe pas de permutation telle que la matrice soit triangulaire par bloc avec des matrices carrées en diagonale). On la prendra aussi stochastique en colonne par analogie avec la théorie des probabilités (chaîne de Markov).

$H$  est déjà à coefficients positifs, il suffit alors seulement de la rendre primitive. Pour cela on ajoute un multiple d'une matrice qui ne contient que des 1. La somme est immédiatement irréductible. Il est intéressant de noter que le graphe associé à une matrice irréductible, positive est fortement connexe. Avec les hypothèses de stochasticté, il s'agit même d'une chaîne de Markov où  $I$  peut être vue comme une loi de probabilité sur le graphe.

**Définition 2** (*Matrice de Google*) : En posant :  $G = \alpha H + \frac{1-\alpha}{N}F$  avec  $F$  une matrice qui ne comporte que des 1, on peut appliquer le théorème de Perron-Frobenius et obtenir le résultat par approximations successives.  $G$  est appelée "matrice de Google"

Cette méthode converge en  $O(N)$ , ce qui est assez élevé pour des graphes de la dimension du graphe de Google.

### 3.2 PageRank Quantique

En 2014 (à vérifier), on propose la première version du PageRank Quantique qui repose sur l'algorithme de Grover. Cette méthode a une complexité en  $O(\sqrt{N})$ .

On définit le système des états comme étant le graphe du Web (espace de Hilbert) dont la matrice d'adjacence est le graphe de Google. L'état  $|i\rangle$  est l'état propre correspondant au  $i^{eme}$  noeud du graphe.

Comme le graphe est orienté, on raisonne sur le produit tensoriel des espaces de Hilbert. Donc,  $|i\rangle \otimes |j\rangle$  représente la liaison  $i \rightarrow j$  du graphe.

On met le système dans un état initial superposé :

$$|\psi_0\rangle = \sum_{j=1}^N (|j\rangle \otimes \sum_{k=1}^N \sqrt{G_{kj}} |k\rangle)$$

Que l'on fait évoluer par réflexions successives autour de l'espace engendré par les  $|\psi_j\rangle = |j\rangle \otimes \sum_{k=1}^N \sqrt{G_{kj}} |k\rangle$ .

**Définition 3** L'opérateur d'évolution est défini de la manière suivante :

$$U = SR = S(2P - I)$$

Avec  $S$  l'opérateur de Swap ( $\sum_{i,j} |ij\rangle \langle ji|$ ) et  $P$  la projection ( $\sum_{i,j} |\psi_j\rangle \langle \psi_i|$ )

On appliquera  $U^2$  à chaque itération pour retrouver des états bien orientés.

Cet algorithme de Grover est en  $O(\sqrt{n})$  et il a été prouvé qu'il n'était pas possible de faire mieux en algorithme de recherche d'un élément parmi un ensemble d'éléments.

Ainsi, il ne sera pas possible de chercher dans cette direction pour mettre en place un nouvel algorithme quantique.

### 3.3 Points de similitudes entre versions classiques et quantiques.

Les versions classique et quantique du PageRank se comportent différemment à la simulation : cela est normal car on n'applique pas exactement la même méthode dans chacun des cas.

Le cas quantique semblerait discriminer moins sévèrement les pages en appliquant des scores plus homogènes. L'algorithme classique aurait tendance à séparer plus sévèrement les valeurs extrêmes mais apporte peu de pertinence dans le choix de pages d'importance semblable.

Une différence fondamentale réside dans le fait que dans le cas quantique, on mesure la valeur moyenne que prend un état dans le temps (les systèmes quantiques itératifs sont rarement stables à l'instant  $t$ ).

La convergence se fait en moyenne dans le temps : les valeurs sont plus lissées et les hubs secondaires (clusters non principaux) revêtent plus d'importance dans le cas Quantique (voir Figure 2) : cela est une interprétation éventuelle du phénomène observé.

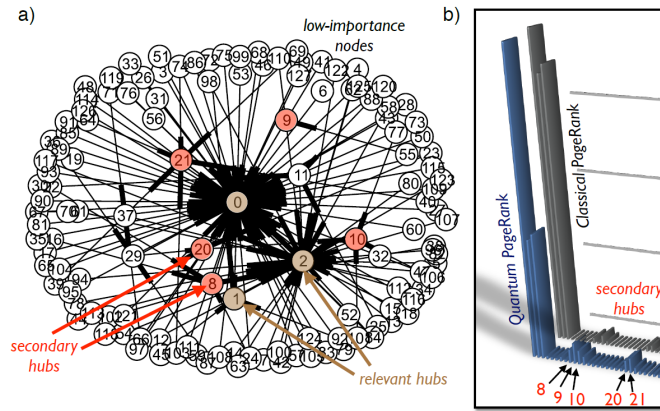


FIGURE 2 – Comparaison des algorithmes classiques et quantiques pour un graphe donné. Voir [5]

Nous nous sommes ainsi reposé là dessus pour trouver une nouvelle voie pour l'algorithme Quantique. Nous avons ainsi observé que l'algorithme de PageRank cherchait à obtenir un résultat comparable à celui que l'on pouvait observer avec une transformée de Fourier : on veut quantifier l'importance relative de certains noeuds, comme la transformée de Fourier met en exergue l'importance relative de certaines fréquences dans le développement d'une fonction sur la base des exponentielles. Cela demande, bien évidemment, des approfondissements tant sur le plan théorique qu'applicatif - cependant, nous avons souhaité retracer le cheminement réel de notre pensée.

Cette observation est appuyée par le fait qu'affiner la détermination des hubs secondaires de PageRank serait rendue plus aisée par une telle méthode car la contribution des harmoniques d'ordre supérieur serait un facteur discriminant pour séparer des pages qui possèdent une importance similaire.

## 4 Apport de la théorie spectrale des Graphes

### 4.1 Contextualisation

La théorie spectrale des graphes est née dans les années 2010 suite au besoin de généraliser la transformée de Fourier aux Graphes. Cette théorie est très riche et prometteuse ; elle trouve des applications dans de nombreux domaines - du machine learning à l'étude de la topologie des graphes. Cette théorie a pris forme pour des graphes non dirigés, très récemment (2019), la théorie a été étendue aux graphes dirigés. Nous allons prendre appui sur cette formalisation de la théorie.

### 4.2 Introduction à la théorie

La théorie spectrale des graphes prend racine dans les origines de la transformée de Fourier : déterminer les modes des solutions de l'équation de la chaleur. On cherche à résoudre l'équation de la chaleur pour un graphe :

$$\partial_t \phi = -\Delta \phi$$

Avec  $\phi$  un signal sur le graphe (ie une fonction de  $\mathbb{R}^N$  avec  $N$  le nombre de noeuds du graphe) et  $\Delta$  un opérateur sur les signaux à définir.  $\phi$  peut être vue comme une distribution de chaleur sur un graphe.

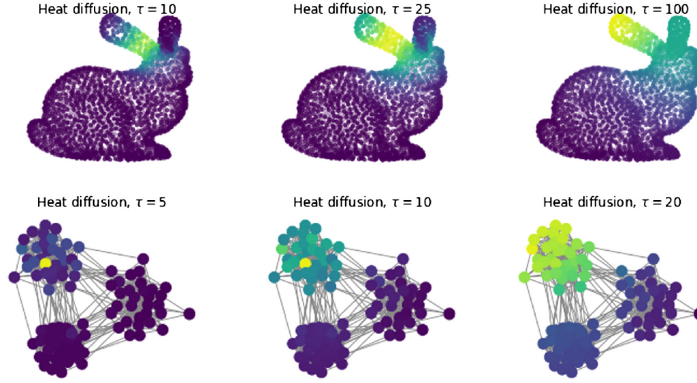


FIGURE 3 – Diffusion de chaleur sur un graphe. Tiré de [2]

On peut retrouver l'expression de  $\Delta$  en utilisant la loi de Newton pour le transfert de chaleur : la chaleur transférée entre les noeuds  $i$  et  $j$  est proportionnelle à  $\phi_i - \phi_j$  si  $i$  et  $j$  sont connectés. C'est à dire, avec  $k$  la capacité calorifique et  $A$  la matrice d'adjacence du Graphe :

$$\begin{aligned}
 \partial_t \phi &= -k \sum_j A_{ij} (\phi_i - \phi_j) \\
 &= -k (\phi_i \deg(\text{noeud}_i) - \sum_j A_{i,j} \phi_j) \\
 &= -k \sum_j (\delta_{i,j} \deg(\text{noeud}_i) - A_{i,j}) \phi_j \\
 &= -k \sum_j l_{i,j} \phi_j \\
 &= -k L \phi
 \end{aligned}$$

On trouve :

**Définition 4** Le Laplacien pour un graphe est défini comme étant l'application suivante, avec  $f$  un signal sur un graphe :

$$\begin{aligned}
 \Delta(f)_{i,j} &= L_{i,j} = -\delta_{i,j} \deg(\text{noeud}_i) + A_{i,j} \\
 \Delta(f) &= L = -\text{diag}(\deg(\text{noeud}_i)) + A
 \end{aligned}$$

Une fois que l'on a défini le Laplacien, on définit le gradient de la manière (naturelle) suivante :

**Définition 5** Le gradient pour un graphe est défini comme étant l'application suivante, avec  $f$  un signal sur un graphe :

$$\nabla f(i, j) = \sqrt{A_{i,j}} (f(j) - f(i))$$

On a bien  $\nabla^T \nabla = \Delta$

L'intérêt du gradient se retrouvera après avoir défini la transformée de Fourier sur un graphe

### 4.3 Transformée de Fourier pour un graphe non orienté

**Définition 6** (*Transformée de Fourier sur un graphe orientée*) : Soit  $u_k$  une base de vecteurs propres du Laplacien d'un graphe  $G=(V,E)$ . La transformée de Fourier sur  $G$  est l'application qui à un signal  $f$  associe  $\hat{f}$  :

$$\hat{f} = \sum_{i \in V} f(i) u_k(i) = \langle f | u_k \rangle$$

On obtient la transformée inverse directement par changement de base :

$$f = \sum_{k=0}^{N-1} \hat{f}(i) u_k(i)$$

Par le théorème du minimax on obtient les valeurs propres en utilisant l'énergie de Dirichlet.

**Propriété-Définition 1** (*Localisation des valeurs propres*) : L'énergie de Dirichlet pour le Laplacien  $L$  et le signal  $f$  est définie par  $E(f) = \langle f | Lf \rangle$ .

On a :

$$E(f) = \langle f | Lf \rangle = \sum_{(i,j) \in E} A_{ij} (f(j) - f(i))^2 = \|\nabla f\|_2^2$$

Immédiatement, si  $f$  est un vecteur propre de  $L$  (disons  $u_i$ ) associé à  $\lambda_i$  :

$$\lambda_i = \frac{E(u_i)}{\|u_i\|_2^2} = \frac{\|\nabla u_i\|_2^2}{\|u_i\|_2^2}$$

Cette propriété est fondamentale pour interpréter la transformée de Fourier pour un graphe : les valeurs propres du Laplacien, à la manière des fréquences, traduisent l'oscillation des vecteurs propres associés. Un vecteur propre dont la norme du gradient (et par extension l'oscillation) est faible sera associé à une valeur propre faible (que l'on peut voir comme une fréquence d'oscillation). On remarque également que la notion d'énergie se rapporte bien à ce qu'on a l'habitude de connaître car la valeur de l'énergie dépend directement de l'oscillation d'un signal sur un graphe.

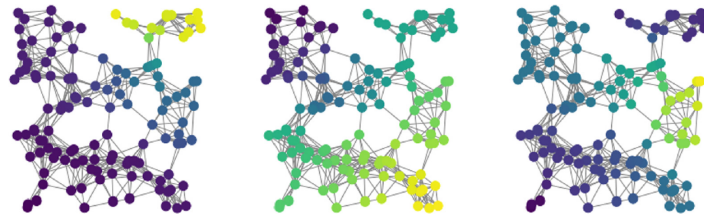


FIGURE 4 – Affichage des vecteurs propres associés aux 3 valeurs propres de plus basse énergie. Tiré de [2]

Ce travail a été effectué jusque-là pour des graphes non-orientés. Nous nous intéressons désormais à une généralisation pour les graphes orientés proposée très récemment (inscrire référence)

## 4.4 Généralisation aux graphes orientés

Nous nous appuyons ici entièrement sur les travaux de HARRY SEVI, GABRIEL RILLING et PIERRE BORGNAT (Février 2019, voir [6]).

La grande particularité des graphes non-orientés à poids réels est que leur matrice d'adjacence est symétrique réelle, ce qui permet de toujours diagonaliser le laplacien dans  $\mathbb{R}$  (théorème spectral). Cela n'est plus le cas pour les graphes orientés et il faut trouver un moyen de contourner cela : la solution réside dans les marches aléatoires.

**Définition 7** (*Opérateur de marche aléatoire*) : On considère une marche aléatoire sur un graphe dirigé fortement connexe  $G = (V, E, w)$  (chaîne de Markov homogène à espace d'états fini et dont les probabilités de transitions sont proportionnelles aux poids des arêtes).

Pour un graphe dirigé la matrice de probabilité de transition est définie par :

$$P_{i,j} = \mathbb{P}(X_{n+1} = j | X_n = i) = (D^{-1}A)_{i,j}$$

Avec  $D$  la matrice diagonale des degrés sortants des noeuds et  $A$  la matrice d'adjacence du graphe.

On introduit le concept d'ergodicité d'une marche aléatoire sur un graphe.

**Définition 8** (*Ergodicité d'une marche aléatoire*) : Une marche aléatoire est dite ergodique si elle est irréductible et apériodique.

**Définition 9** (*Irréductibilité d'une marche aléatoire*) : Une marche aléatoire est irréductible si la probabilité d'atteindre un noeud  $x$  depuis un noeud  $y$  en un temps fini est strictement positive. On a vu que cela était équivalent à avoir un graphe fortement connexe. La matrice de google est irréductible.

**Définition 10** (*Apériodicité d'une marche aléatoire*) : Une marche aléatoire est apériodique si :

$$\forall x \in V, \rho(x) = \text{pgcd}\{n \in \mathbb{N}^* : P^n(x, x) > 0\} = 1.$$

Cela correspond à une marche aléatoire où l'on ne peut pas revenir sur un noeud pour un multiple d'une période.

La notion d'ergodicité est utile car si la matrice de transition  $\mathbb{P}$  est diagonalisable, le théorème de Perron Frobenius s'applique.

Par ailleurs, une marche aléatoire ergodique possède vers une distribution stationnaire de probabilité (ie  $(P^n)$  converge).

La notion de réversibilité est également intéressante car elle nous permet de quantifier la propriété d'être dirigé pour un graphe.

**Définition 11** (*Réversibilité d'une marche aléatoire*) : La marche aléatoire renversée pour une marche aléatoire ergodique est définie par la matrice de transition suivante :

$$P_{i,j}^* = \frac{\pi_i}{\pi_j} \times P_{j,i}$$

$$P^* = \Pi^{-1} P^T \Pi$$

Avec  $\pi$  la distribution stationnaire de probabilité.

Une marche aléatoire est réversible si  $P = P^*$



Les marches aléatoires sur des graphes non-dirigés sont réversibles, l'inverse n'est pas vrai. On peut rendre la marche aléatoire du Pagerank ergodique et réversible :

**Propriété-Définition 2** (*Lazy Random Walk*) On définit la marche aléatoire apériodique suivante :

$$\tilde{P} = \frac{Id + P}{2}$$

*Il est facilement vérifiable que cette marche aléatoire est apériodique : modifier la matrice de transition de la sorte revient à ajouter des arcs bouclés au graphe de départ.*

*On ne considèrera donc, désormais, que des marches aléatoires ergodiques*

*(Additive reversibilization) On rend la matrice de transition d'une marche aléatoire ergodique réversible :*

$$\bar{P} = \frac{P + P^*}{2}$$

La marche aléatoire obtenue étant ergodique réversible, on définit le Laplacien (qui est une généralisation du Laplacien ci-dessus) pour le graphe.

**Définition 12** (*Laplacien d'un graphe orienté*) : Le Laplacien du graphe est :

$$L = \Pi L_{RW}$$

On a noté  $L_{RW}$  le laplacien de la marche aléatoire.

$$L_{RW} = I - \bar{P}$$

*On vérifiera facilement que ces propriétés sont des généralisations de propriétés obtenues pour les graphes non-orientés*

L'on retrouve les notions d'Energie de Dirichlet définies ci-dessus. On peut définir le gradient d'une façon similaire et on retrouve les idées d'oscillation présentées ci-dessus avec les mêmes formules (voir directement les travaux de HARRY SEVI, GABRIEL RILLING et PIERRE BORGNAT [6]). La grande différence est que les valeurs propres de  $P$  ne sont pas nécessairement réelles.

## 5 Mise au point d'un algorithme Quantique de résolution du PageRank.

### 5.1 Présentation Sommaire

Nous ne ferons ici qu'esquisser les principales idées de l'algorithme : nous avons pour ambition de continuer le projet en 2ème année et nous nous attacherons à préciser pour le mieux ce qui ne demeure ici qu'un balayage sommaire du sujet.

L'idée de l'algorithme est d'utiliser la transformée de Fourier sur les graphes pour faire une approximation de PageRank qui mette davantage en valeur les hubs d'importance secondaires tout en demeurant plus efficace en coûts de calculs.

L'algorithme consiste en les étapes suivantes :

- Effectuer une approximation de faible rang de la matrice du Laplacien de la marche aléatoire. Le théorème d'Eckart-Young assure la faisabilité d'une telle approximation.
- Cette approximation s'effectue par SVD ou PCA. Il existe plusieurs algorithmes quantiques qui réalisent ces opérations en  $O(\log(N))$  avec  $N$  la taille de la matrice de départ (voir par exemple Quantum Higher Order Singular Value Decomposition de Lejia Gu, Xiaoqiang Wang, H. W. Joseph Lee, and Guofeng Zhang [7].)
- L'importance résultante sera une somme des vecteurs singuliers obtenus par PCA. Il conviendra de pondérer les vecteurs de fréquence différentes par un coefficient relatif à leurs participation au signal de départ.

Il demeure de nombreuses questions pour la suite du travail.

- Une première barrière demeure de réussir à mettre le système dans un état superposé qui nous permettrait d'obtenir le résultat souhaité. Chaque algorithme quantique de SVD a sa propre manière de procéder, une première tâche est de trouver la manière qui correspond le plus à nos données initiales et qui soit réalisable avec un nombre de Qbits raisonnable (le laplacien est une matrice de taille de l'ordre de  $10^9$ ...)
- Une seconde barrière est de quantifier les résultats de notre algorithme par rapport à l'algorithme classique de PageRank : ici l'idée principale de l'algorithme change et il convient de quantifier les modifications que cela induit.
- Une dernière barrière serait par exemple le choix des coefficients de pondération des coefficients qui ne peut être arbitraire. Quel jeu de coefficients maximise la pertinence des résultats sans nuire aux hubs secondaires ?

## 6 Conclusion et remerciements

Ce projet passionnant clôt notre première année d'études à CentraleSupélec. Ce fut une expérience profondément enrichissante : sans cesse en proie à la difficulté, nous poursuivions nos efforts pour arriver à des résultats concluants en fin de parcours.

Pour finir, je souhaite remercier tout particulièrement Zeno Toffano et Benoit Valiron, chargés de l'encadrement de notre projet, pour leur présence et implication tout au long de cette étude.

## 7 Bibliographie

### Références

- [1] D. Bacon, “The quantum fourier transform and jordan’s algorithm.”  
Une présentation complète de la transformée de Fourier Quantique. Permet de déceler les enjeux calculatoires de cette opération.
- [2] N. T. P. G. P. V. Benjamin Ricaud, Pierre Borgnat, “Fourier could be a data scientist : From graph fourier transform to signal processing on graphs,” in *Compte rendus Physique*, Septembre 2019.  
Les auteurs de cet article font état des découvertes récentes en théorie spectrale des graphes et décrivent les enjeux à venir.
- [3] F. Chung., “Laplacians and the cheeger inequality for directed graphs.”  
Calcul de nombreuses propriétés de la transformée de Fourier dans les graphes faisant intervenir les marches aléatoires.
- [4] B. R. David Shuman and P. Vandergheynst, “A windowed graph fourier transform,” 2012.  
Une présentation dans laquelle le concept de Transformée de Fourier en fenêtre pour un graphe est très bien expliquée.
- [5] F. C. G.D. Paparo ; M. Muller and M. Martin-Delgado, 2014.
- [6] P. B. Harry Sevi, Gabriel Rilling, “Harmonic analysis on directed graphs and applications : From fourier analysis to wavelets.” 2019.  
Première formalisation de la théorie spectrale des Graphes appliquée aux graphes orientés.
- [7] H. W. J. L. Lejia Gu, Xiaoqiang Wang and G. Zhang, “Quantum higher order singular value decomposition,” April 2020.  
Une présentation de la SVD quantique.
- [8] M. M. Patrick Rebentrost and S. Lloyd, “Quantum support vector machine for big data classification,” *ArXiv*, Juillet 2014.  
Utilisation du calcul quantique dans le calcul de SVM (big data) en avec une complexité logarithmique
- [9] M. Tepper and G. Sapiro, “A short-graph fourier transform via personalized pagerank vectors,” in *IEEE*, 2016, pp. 4806–4810.  
Dans cet article, les auteurs fournissent une application de la transformée de fourier à un graph pour le calcul de vecteurs de pageRank personnalisés.