

Approximation de l'entropie de von Neumann de graphes pour une analyse de vulnérabilité

Tristan AVERTY, Delphine DARÉ-EMZIVAT, Abdel-Ouahab BOUDRAA, Yves PRÉAUX

IRENav, École Navale/Arts-Metiers ParisTech, CC 600, 29240 Brest Cedex 9, France.

[tristan.averty, delphine.dare, abdel.boudraa, yves.preaux]@ecole-navale.fr

Résumé – Dans ce travail, nous exploitons la variation de l'entropie de von Neumann de graphes comme mesure de vulnérabilité en proposant une nouvelle forme approchée de cette entropie basée sur des attributs structurels du graphe, à savoir le nombre d'arêtes, de sommets ou encore des degrés du graphe. L'utilisation d'une telle forme est motivée par l'optimisation du temps de calcul qui en découle. Disposant d'une forme simplifiée de l'entropie, nous l'utilisons pour la caractérisation de la vulnérabilité des graphes via l'étude de la variation entropique du graphe suite à la suppression d'arêtes. Les résultats obtenus sur des graphes de grandes tailles montrent la pertinence d'une telle approximation.

Abstract – In this work, we use the von Neumann graph entropy variation as a measure of graph vulnerability and we propose a new approximate form of this entropy based on structural attributes of the graph, namely the number of edges, vertices or degrees of the graph. The use of such a simplification is mainly motivated by the resulting optimized computation time. Having a simplified form of the entropy, we use it as a characterization of graph vulnerability through the study of the entropy variation following the deletion of edges. Obtained results in large size graphs show the effectiveness of the proposed approximation.

1 Introduction

Nos sociétés dépendent de plus en plus d'infrastructures qui interagissent entre elles de manière collaborative et coordonnée afin de faciliter l'interaction humaine et fournir des services essentiels. Qu'il s'agisse des installations ferroviaires, des connexions internet, des réseaux de transmission de gaz ou encore des relations humaines dans une communauté, tous ces aspects de la société peuvent être caractérisés par leur fonctionnement en réseau. Ces systèmes peuvent être représentés sous la forme de graphe et nous disposons alors de l'architecture du réseau définissant les liaisons existantes entre les différents composants du réseau et la hiérarchie éventuelle entre eux. Dans un travail de recherche précédent, nous avons développé une stratégie de mesure de la vulnérabilité des connexions d'un réseau modélisé par un graphe, celui-ci étant vu comme un système quantique [1],[2]. Nous avons établi que la matrice de densité associée à ce graphe à la suite d'une perturbation du poids d'une arête agit sur l'entropie de von Neumann (VNGE pour *von Neumann Graph Entropy*) résultante. L'algorithme ainsi développé et les résultats obtenus ont confirmé la pertinence de cette nouvelle mesure mais n'est pas applicable en l'état sur de grands graphes en raison de sa complexité élevée ($O(n^3)$ où n est le nombre de sommets du graphe considéré) inhérente au calcul du spectre de la matrice laplacienne. Afin d'améliorer l'efficacité de ce calcul, une démarche de simplification et d'approximation a été entreprise avec toujours pour objectif la caractérisation de la vulnérabilité des graphes.

La VNGE d'un réseau a été introduite par Braunstein et al. [3] puis utilisée et analysée en détails dans un certain nombre

de travaux ultérieurs [4],[5],[6]. Pour la plupart d'entre eux, l'idée derrière cette mesure est d'abord de définir la matrice de densité d'un graphe en se basant sur la correspondance entre états quantiques et matrice laplacienne, puis de caractériser ce dernier grâce à cette entropie. Une telle approche fait appel au calcul du spectre laplacien des graphes. La complexité d'un tel calcul est cubique par rapport au nombre de nœuds du réseau, ce qui rend l'application aux grands réseaux irréalisable. Han et al. [6] ont cherché à surmonter ce problème en examinant ce qu'ils nomment l'« entropie quadratique » permettant d'approcher, à l'ordre 2, l'entropie de Shannon d'un graphe (qui n'a de différence avec la VNGE que la base logarithmique). Choi et al. [7] ont, quant à eux, proposé des approximations quadratiques de la VNGE à l'aide de polynômes.

Dans cet article, nous présentons une stratégie d'optimisation du calcul de la VNGE en exploitant deux approximations quadratiques de la fonction $f(x) = x \ln x$: une basée sur un développement limité et l'autre basée sur un polynôme. Le but étant d'optimiser le calcul d'une carte de vulnérabilité [1],[2], des tests sont réalisés sur des réseaux connus de la littérature.

2 Entropie d'un graphe

Soit un graphe $G = (\mathcal{V}, \mathcal{E})$ simple, non orienté et non pondéré, avec $|\mathcal{V}| = n$ sommets et $|\mathcal{E}| = m$ arêtes. La matrice d'adjacence \mathbf{A} du graphe G est une matrice de taille $n \times n$ ayant pour coefficient $A_{ij} = A_{ji} = 1$ lorsque les sommets i et j sont reliés par une arête et $A_{ij} = A_{ji} = 0$ sinon. La matrice des degrés \mathbf{D} , diagonale, admet pour coefficient $D_{ii} = d(i) = \sum_{j=1}^n A_{ij}$ le degré du sommet i . Enfin, la ma-

trice laplacienne est définie par $\mathbf{L} := \mathbf{D} - \mathbf{A}$. Bien que les matrices \mathbf{A} et \mathbf{L} soient toutes les deux utilisées à des fins de représentations de graphe, c'est la matrice laplacienne qui sert à définir l'entropie du graphe G . En effet, tout graphe peut être vu comme un système physique quantique [2] c'est-à-dire possédant une matrice de densité, définie par Braunstein et al. [3] grâce à la matrice laplacienne comme étant $\rho := \mathbf{L}/\text{Tr}(\mathbf{L})$. Dans le but d'accélérer les calculs, les relations suivantes sont vérifiées [7],[8] :

$$\text{Tr}(\rho) = \sum_{\ell=1}^n \mu_\ell = 1, \quad \text{Tr}(\rho^2) = \sum_{\ell=1}^n \mu_\ell^2 = \frac{Z(G)}{(2m)^2} + \frac{1}{2m} \quad (1)$$

où $\mu_1 \geq \dots \geq \mu_n = 0$ sont les valeurs propres de ρ , $Z(G) = \sum_{i \in \mathcal{V}} d(i)^2$ est le « premier indice de Zagreb » du graphe G et où la quantité $\text{Tr}(\rho^2)$ est appelée « pureté » de la matrice de densité ρ . Deux éléments sont importants à noter à partir de ces formules : la première est que la pureté ne dépend que d'attributs structurels du graphe G , à savoir le nombre d'arêtes et les degrés des sommets ; la deuxième est que, en supposant que la matrice \mathbf{D} soit un attribut du graphe créé en même temps que ce dernier, la complexité du calcul de la pureté passe de $O(n^2)$ à $O(n)$.

Lorsque la matrice de densité est définie, la VNGE du graphe G se calcule à partir des valeurs propres de cette matrice [4] :

$$S(G) := S(\rho) = -\text{Tr}(\rho \ln \rho) = -\sum_{\ell=1}^n \mu_\ell \ln \mu_\ell \quad (2)$$

L'égalité $0 \ln 0 = 0$ étant admise [3].

3 Approximations de l'entropie

Le principal problème du calcul de la VNGE est qu'il y a nécessité de calculer toutes les valeurs propres de la matrice ρ . Un tel calcul relève d'une complexité $O(n^3)$. Lorsque n est trop grand (ce qui est le cas dans l'analyse de graphes issus de modèles réels), le temps de calcul augmente donc considérablement. Une manière d'alléger les calculs est de passer par une approximation de l'entropie (2) c'est-à-dire par une approximation de la fonction $f(\mu_\ell) := \mu_\ell \ln \mu_\ell$.

3.1 Développements limités

Inspirés par les travaux de Han et al. [6] qui présentent la fonction $\mu_\ell(1 - \mu_\ell)$ comme étant une approximation de $f(\mu_\ell)$ (on notera $S^H(G)$ l'approximation de la VNGE résultante), une première approximation de la VNGE de G peut être obtenue à l'aide d'un développement limité (DL) à l'ordre 2 de la fonction $f(\mu_\ell)$.

Proposition 1. *Soit un graphe $G = (\mathcal{V}, \mathcal{E})$ avec $|\mathcal{V}| = n$ sommets et $|\mathcal{E}| = m$ arêtes. La VNGE de G est approchée, autour d'un point $0 < \hat{\mu} \leq 1$, par :*

$$S_{\text{DL}, \hat{\mu}}(G) = \frac{n\hat{\mu}}{2} - \ln \hat{\mu} - \frac{1}{2\hat{\mu}} \text{Tr}(\rho^2) \quad (3)$$

Preuve. L'approximation est calculée en remplaçant $\mu_\ell \ln \mu_\ell$ dans (2) par le développement limité autour de $\hat{\mu}$ de la fonction $f(\mu_\ell)$ suivant :

$$\mu_\ell \ln \mu_\ell \sim_{\hat{\mu}} \hat{\mu} \ln \hat{\mu} + (\mu_\ell - \hat{\mu})(\ln \hat{\mu} + 1) + \frac{1}{2\hat{\mu}}(\mu_\ell - \hat{\mu})^2$$

Ainsi,

$$\begin{aligned} S_{\text{DL}, \hat{\mu}}(G) &= -\sum_{\ell=1}^n \hat{\mu} \ln \hat{\mu} + (\mu_\ell - \hat{\mu})(\ln \hat{\mu} + 1) + \frac{1}{2\hat{\mu}}(\mu_\ell - \hat{\mu})^2 \\ &= \frac{n\hat{\mu}}{2} - \ln \hat{\mu} \sum_{\ell=1}^n \mu_\ell - \frac{1}{2\hat{\mu}} \sum_{\ell=1}^n \mu_\ell^2 \\ &\stackrel{(1)}{=} \frac{n\hat{\mu}}{2} - \ln \hat{\mu} - \frac{1}{2\hat{\mu}} \text{Tr}(\rho^2) \quad \blacksquare \end{aligned}$$

La complexité du calcul de $S_{\text{DL}, \hat{\mu}}(G)$ est abaissée, grâce à l'équation (1), à $O(n)$.

La recherche d'un $\hat{\mu}$ optimal est intéressante mais non-triviale. Néanmoins, dans certains cas (en général les graphes possédant la propriété de « petit-monde »), les valeurs propres sont concentrées autour de leur moyenne [7] : $\frac{1}{n}$. L'approximation de l'entropie optimale à l'aide d'un développement limité d'ordre 2 peut donc être $S_{\text{DL}, 1/n}(G)$.

3.2 Approximation polynomiale

Le problème de l'approximation précédente est le choix laissé de la valeur $\hat{\mu}$. Il peut être plus naturel d'essayer d'approcher $f(\mu_\ell)$ par un polynôme d'ordre 2 avec des contraintes adaptées. Des travaux dans ce sens ont été menés par Choi et al. [7] et on notera $S^C(G)$ l'approximation de la VNGE résultante. Le problème de cette dernière est qu'elle fait intervenir la valeur propre maximale μ_1 qui, bien que moins coûteuse à calculer que le spectre entier grâce à la méthode de la puissance itérée [9], impose toutefois un calcul de valeur propre. Pour passer outre ce calcul, une solution est d'approcher la fonction $f(\mu_\ell) = \mu_\ell \ln \mu_\ell$ par un polynôme $q(\mu_\ell) = a\mu_\ell^2 + b\mu_\ell + c$ de degré 2 vérifiant les contraintes $q(\frac{1}{n}) = f(\frac{1}{n})$, $q'(\frac{1}{n}) = f'(\frac{1}{n})$ et $q(1) = f(1)$. Le choix a été fait de mettre deux contraintes en $\frac{1}{n}$, moyenne des valeurs propres, pour les mêmes raisons que précédemment.

Proposition 2. *Soit un graphe $G = (\mathcal{V}, \mathcal{E})$ avec $|\mathcal{V}| = n$ sommets et $|\mathcal{E}| = m$ arêtes. La VNGE de G est approchée par :*

$$S_{\text{poly}}(G) = -(a \text{Tr}(\rho^2) + b + nc) \quad (4)$$

avec

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \frac{1}{(n-1)^2} \begin{pmatrix} n^2(\ln n - 1) + n \\ n^2(1 - \ln n) - 1 - \ln n \\ -n + 1 + \ln n \end{pmatrix}$$

Preuve. L'approximation est calculée en remplaçant, de la même manière que dans la démonstration de la Proposition 1, l'expression $\mu_\ell \ln \mu_\ell$ dans (2) par le polynôme $q(\mu_\ell) = a\mu_\ell^2 + b\mu_\ell + c$ où les valeurs de a, b et c sont obtenues en résolvant le système linéaire associée aux contraintes, c'est-à-dire le système :

$$\begin{cases} a + bn + cn^2 = -n \ln n \\ 2a + bn = n(1 - \ln n) \\ a + b + c = 0 \end{cases} \quad \blacksquare$$

Les deux approximations de $f(x) = x \ln x$ de la littérature ainsi que les deux proposées dans cet article sont présentées en figure 1 en supposant que le graphe étudié possède 10 sommets et que $\mu_1 = 0.6$. Par construction, nos deux approximations sont tangentes à la courbe $f(x)$ au point $x = \frac{1}{n}$ et le polynôme passe bien par le point $(1, 0)$.

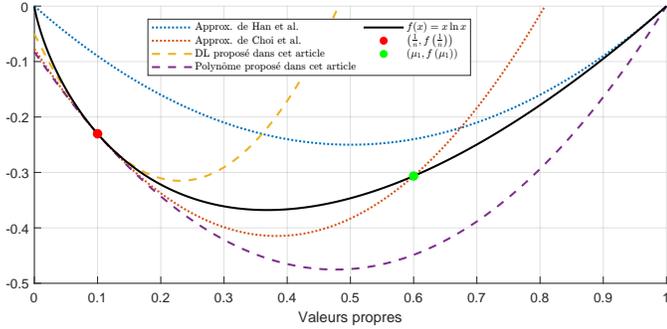


FIGURE 1 – Représentation des approximations d’ordre 2 de la fonction $x \ln x$ (deux issues de la littérature ainsi que les deux présentées dans cet article) en supposant $n = 10$ et $\mu_1 = 0.6$.

3.3 Erreur d’approximation et temps de calcul

Dans le but de montrer la plus-value des entropies approchées, nous étudions l’évolution des temps nécessaires à leurs calculs ainsi que les erreurs commises par rapport à la VNGE, et ce, uniquement en fonction du nombre de sommets du graphe car les différents calculs ne dépendent pas du nombre d’arêtes.

La figure 2 présente les temps de calcul (moyennés sur 10 itérations) de nos deux entropies approchées, des deux approximations de la littérature ainsi que de l’entropie initiale, toutes calculées à partir de graphes aléatoires suivant le modèle de Watts-Strogatz (graphes possédant la propriété de petit-monde) avec un nombre de sommets croissant. Il est facile de voir sur la figure 2 que le calcul de la VNGE devient rapidement inconfortable lorsque le nombre de sommets augmente, cette dernière ayant une complexité en $O(n^3)$. L’approximation $S^C(G)$ de Choi et al., bien que plus rapide à calculer que l’entropie exacte, est tout de même moins intéressante que les trois autres à cause du temps nécessaire au calcul de μ_1 . En effet, la figure 2 montre que nos deux approximations ainsi que celle de Han et al. nécessitent un temps de calcul beaucoup plus faible, surtout pour un nombre conséquent de sommets, ce qui rend leur utilisation intéressante pour traiter de grands graphes.

Dans le but d’utiliser ces approximations, il faut également que la différence entre la VNGE et les versions approchées ne soit pas trop grande et soit décroissante en fonction du nombre de sommets : cette assertion est vérifiée pour trois des quatre approximations. En effet, la figure 2 montre des courbes décroissantes d’erreurs absolues moyennes (moyennées sur 10

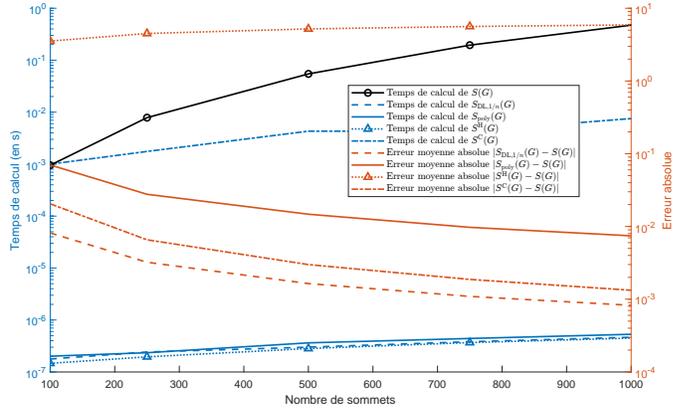


FIGURE 2 – Représentation des temps de calcul (en bleu) et des erreurs absolues (en orange) pour $S(G)$, $S_{DL,1/n}(G)$, $S_{poly}(G)$, $S^H(G)$ et $S^C(G)$.

itérations) en fonction du nombre de sommets, avec un avantage pour notre approximation issue d’un développement limité. Par ailleurs, il est clair que l’approximation de Han et al. est moins performante en termes d’erreurs car l’« entropie quadratique » qu’ils utilisent sur-estime assez largement la fonction $f(\mu_\ell)$ sur tout l’intervalle $[0, 1]$ (figure 1) indépendamment du nombre de sommets. Une conséquence est l’obtention d’une erreur très importante sur l’approximation de la VNGE.

4 Vulnérabilité de graphes

La vulnérabilité peut être appréhendée d’un point de vue défensif ou offensif. Latora et al. [10] définissent les arêtes (ou nœuds) les plus sensibles qui correspondent à ceux qui conduisent à l’obtention de la valeur la plus préjudiciable pour leur grandeur « Performance » du graphe. Chen et al. [11] définissent la vulnérabilité comme la capacité d’un graphe à résister à une attaque virale et assimile la vulnérabilité d’un graphe donné à la valeur maximale du module des valeurs propres de \mathbf{A} . Nous avons développé une méthode caractérisant cette vulnérabilité appelée VPV (pour von Neumann Perturbation Vulnerability) à partir de la variation d’entropie liée à des suppressions d’arêtes [1],[2]. Il est à noter que la relation entre vulnérabilité d’un réseau et entropie n’est pas clairement formalisée à ce jour. Cet algorithme est constitué des étapes suivantes :

1. Calcul de l’entropie $M(G)$ du graphe initial $G = (\mathcal{V}, \mathcal{E})$.
2. Pour chaque arête, calcul de la variation de l’entropie si cette arête est supprimée et stockage de cette variation dans une matrice \mathbf{W}_M :

$$(W_M)_{ij} = \frac{M(\tilde{G}) - M(G)}{M(G)} \quad (5)$$

où $\tilde{G} = (\mathcal{V}, \mathcal{E} \setminus \{(i, j)\})$ est le graphe initial auquel l’arête (i, j) a été enlevée.

3. Création du graphe pondéré $G_M = (\mathcal{V}, \mathcal{E}, \mathbf{W}_M)$.

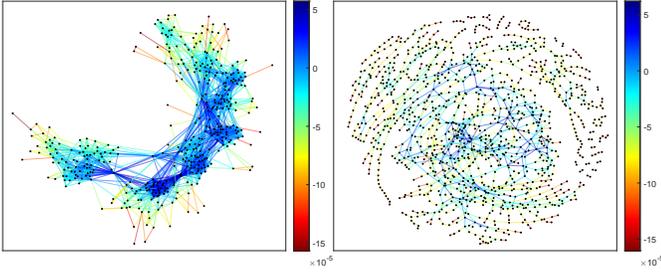


FIGURE 3 – Analyse VPV (entropie exacte) de deux graphes connus de la littérature : Infect Dublin (à gauche) / Euro Road (à droite)

La figure 3 montre l'analyse VPV effectuée sur deux graphes de la littérature [12] avec $M(\cdot) = S(\cdot)$. Le premier réseau considéré est le réseau de contacts humains Infect Dublin (410 nœuds / 2765 arêtes) issu d'une collecte de données durant l'événement *Infectious SocioPatterns* à la galerie des sciences de Dublin. Les nœuds représentent des personnes et les arêtes les contacts entre ces personnes. Le deuxième est le réseau routier Euro Road (1174 nœuds / 1417 arêtes). Les nœuds représentent des villes et les arêtes représentent des routes entre ces villes. La couleur d'une arête correspond à l'évolution entropique issue de sa suppression dans le graphe : plus l'arête est rouge, plus elle est responsable d'une perte d'entropie et est interprétée comme vulnérable. Le tableau 1 résume les performances des temps de calcul et les erreurs moyennes des analyses VPV lorsque les entropies approchées sont considérées. Il est clair que lorsque la fonction $M(\cdot)$ de l'algorithme VPV est la VNGE $S(\cdot)$, le temps de calcul augmente considérablement lorsque le nombre de sommets augmente. Cette remarque est également valable lorsque $M(\cdot) = S^C(\cdot)$. Pour les trois autres analyses VPV, les temps de calcul sont inférieurs à 1 s avec un facteur de réduction de près de 147 pour le cas du graphe Infect Dublin et un facteur de réduction de 3767 dans le cadre du second. L'analyse de graphes réels est donc plus confortable en considérant les entropies approchées. La caractérisation de l'erreur moyenne absolue, décrite par la formule

$$E(G_{M_1}, G_{M_2}) = \frac{1}{m} \sum_{(i,j) \in \mathcal{E}} \left| (W_{M_1})_{ij} - (W_{M_2})_{ij} \right| \quad (6)$$

est également intégrée dans le tableau 1. L'erreur diminue lorsque le nombre de nœuds augmente ce qui était déjà un point relevé lors de l'analyse de la figure 2. Il est à noter que l'approximation issue du DL de $f(\mu_\ell)$ est la plus performante en termes d'erreurs sur les deux graphes, dû au fait que cette approximation est plus correcte autour de la moyenne du spectre.

5 Conclusion

Dans ce travail, nous avons procédé à deux approximations de la VNGE s'émancipant d'un quelconque calcul de valeurs propres (spectre entier ou valeur propre maximale) afin d'accélérer le calcul de cette entropie notamment pour des grands graphes : l'une portant sur un développement limité d'ordre 2,

		Infect Dublin	Euro Road
Temps (en s)	G_S	79.34	1066.14
	G_{S^H}	0.565	0.282
	$G_{S^{DL,1/n}}$	0.544	0.283
	G_{S^C}	3.54	2.838
	$G_{S^{poly}}$	0.549	0.283
Erreurs	$E(G_S, G_{S^H})$	2.18×10^{-5}	5.48×10^{-5}
	$E(G_S, G_{S^{DL,1/n}})$	0.91×10^{-5}	1.27×10^{-5}
	$E(G_S, G_{S^C})$	1.21×10^{-5}	4.42×10^{-5}
	$E(G_S, G_{S^{poly}})$	2.19×10^{-5}	5.49×10^{-5}

TABLEAU 1 – Tableau intégrant les temps de calcul et les erreurs moyennes absolues des différentes analyses VPV pour deux graphes de la littérature.

la seconde sur une approximation polynomiale quadratique.

Nous avons démontré la pertinence de notre démarche avec un calcul d'une carte de vulnérabilité des arêtes d'un graphe construite à partir de la variation entropique engendrée suite à leurs suppressions. Ainsi nous avons observé la réduction d'un facteur 3767 du temps de calcul sur un graphe comportant 1174 nœuds et 1417 arêtes pour une erreur moyenne absolue avec l'analyse issue de l'entropie exacte égale à 1.27×10^{-5} .

Les approximations présentées dans cet article sont d'ordre 2 dans le but d'accélérer un maximum les calculs. Ce faisant, $\text{Tr}(\rho^2)$ est calculée de manière exacte. Des approximations polynomiales d'ordre supérieur sont à l'étude en ayant à l'esprit que $\text{Tr}(\rho^k)$ peut être approchée par des méthodes stochastiques comme la SLQ (pour *Stochastic Lanczos Quadrature*) [13].

Par ailleurs, une analyse plus approfondie est nécessaire pour formaliser le rôle que joue l'évolution entropique en tant que mesure de vulnérabilité d'une arête. Par exemple, il semble que la suppression de certaines arêtes soient responsables d'une augmentation de l'entropie (figure 3) sans qu'aucune justification n'ait été apportée à ce jour.

Références

- [1] H. A. Bay-Ahmed. *Classification des signaux et des graphes par approches spectrales algébriques*. Thèse de doctorat, EDSML, UBO, Brest, 2018.
- [2] H. A. Bay-Ahmed, D. Daré-Emzivat, and A. Boudraa. Analyse de la vulnérabilité d'un réseau via la mesure de l'entropie de Von Neumann. *GRETSI*, 2019.
- [3] S. Braunstein et al. The Laplacian of a graph as a density matrix : a basic combinatorial approach to separability of mixed states. *Ann. Comb.*, 10(3) :291–317, 2006.
- [4] F. Passerini and S. Severini. The von Neumann entropy of networks. *arXiv :0812.2597*, 2008.
- [5] K. Anand et al. Shannon and von Neumann entropy of random networks with heterogeneous expected degree. *Phys. Rev. E*, 83(3) :036109, 2011.
- [6] L. Han et al. Graph characterizations from von Neumann entropy. *Pattern Recognition Letters*, 33(15) :1958–1967, 2012.
- [7] H. Choi et al. Fast computation of von Neumann entropy for large-scale graphs via quadratic approximations. *Linear Algebra Appl.*, 585 :127–146, 2020.
- [8] I. Milovanovic et al. Remarks on the sum of powers of laplacian eigenvalues of graphs. *the International Mathematical Virtual Institute*, 10(3) :437–447, 2020.
- [9] R. von Mises and H. Pollaczek-Geiringer. Praktische Verfahren der Gleichungsauflösung. *Journal of Applied Mathematics and Mechanics*, 9(1) :58–77, 1929.
- [10] V. Latora and M. Marchiori. Vulnerability and protection of infrastructure networks. *Physical Review E*, 71(1) :015103, 2005.
- [11] C. Chen et al. Node immunization on large graphs : Theory and algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 28(1) :113–126, 2015.
- [12] Ryan A. Rossi and Nesreen K. Ahmed. The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI*, 2015.
- [13] S. Ubaru et al. Fast estimation of $\text{Tr}(f(A))$ via stochastic lanczos quadrature. *SIAM Journal on Matrix Analysis and Applications*, 38(4) :1075–1099, 2017.