

Caractérisation quantitative des réseaux de fibronectine dans la matrice extracellulaire tumorale : vers des biomarqueurs pronostiques et prédictifs de réponse à l’immunothérapie

Faisal JAYOUSI¹ Xavier DESCOMBES¹ Emmanuel BOUILHOL² Ellen VAN OBERGHEN-SCHILLING² Laure BLANC-FÉRAUD¹

¹Laboratoire d’Informatique, Signaux et Systèmes (I3S), 2000 Rte des Lucioles, 06900 Sophia Antipolis, France

²Institut de Biologie Valrose (IBV), Université Côte d’Azur, Parc Valrose, 06108 Nice, France

Résumé – La matrice extracellulaire (MEC) est un réseau de protéines qui jouent un rôle crucial dans la régulation du comportement cellulaire, de la signalisation et de l’architecture tissulaire. Cette étude propose un cadre computationnel pour caractériser quantitativement l’organisation fibrillaire de la fibronectine (FN), un des composants clé de la MEC tumorale, en utilisant des mesures fondées sur les graphes. Nous analysons d’une part les images confocales de la FN assemblée par des cultures de fibroblastes, traitées pour mimer des phénotypes des fibroblastes associés aux tumeurs, et d’autre part des images multispectrales de la FN dans les échantillons de tissu tumoral. Notre méthode consiste à réhausser les images de FN à l’aide de filtres de Gabor, à extraire une représentation squelettique des réseaux fibrillaires et en déduire un graphe pour en capturer la connectivité et les propriétés géométriques. Des caractéristiques clés telles que l’alignement des fibres, la tortuosité et les mesures de centralité sont calculées pour décrire la structure de la FN. Ces caractéristiques sont utilisées pour classer les assemblages de FN en catégories de type tumeur et de contrôle, atteignant un F1-score de 90% avec une forêt d’arbres décisionnels, surpassant un modèle d’apprentissage profond (EfficientNet, 85.71%). L’approche proposée est validée sur des données *in vitro* et appliquée avec succès à des images cliniques, démontrant sa capacité à segmenter et à caractériser les réseaux de FN dans des microenvironnements tumoraux complexes.

Abstract – The extracellular matrix (ECM) is a network of proteins that play a critical role in regulating cellular behaviour, signalling, and tissue architecture. This study proposes a computational framework to quantitatively characterise the fibrillar organisation of fibronectin (FN), a key component of the tumour ECM, using graph-based and statistical measures. We analysed both confocal images of FN assembled by cultured fibroblasts, treated to mimic the phenotypes of cancer-associated fibroblasts, and multispectral images of FN in samples of tumour tissue. Our method involves enhancing FN images using Gabor filters, extracting skeletal representations of fibrillar networks, and constructing graphs to capture connectivity and geometric properties. Key features such as fibre alignment, tortuosity, and centrality measures are computed to describe the FN structure. These features are used to classify FN assemblies into tumour-like and control categories, achieving an F1-score of 90.00% with a Random Forest classifier, outperforming a state-of-the-art deep learning model (EfficientNet, 85.71%). The proposed approach is validated on *in vitro* data and successfully applied to clinical images, demonstrating its ability to segment and characterise FN networks in complex tumour microenvironments.

1 Introduction

Les cellules évoluent dans un microenvironnement complexe et dynamique qui régule leur comportement, leur signalisation, leur prolifération et leur différenciation. Ce microenvironnement est structuré par la matrice extracellulaire (MEC), un réseau tridimensionnel de protéines et de glycanes qui à la fois assure un support mécanique et transmet des signaux biochimiques essentiels aux cellules environnantes.

La fibronectine (FN) est l’un des composants de la MEC. Il s’agit d’une glycoprotéine fibrillaire multifonctionnelle jouant un rôle important dans les processus développementaux et pathologiques.

L’organisation géométrique de la FN au sein de la MEC gouverne les fonctions cellulaires. L’alignement des fibres, à titre d’exemple, régule la migration des cellules environnantes, et de ce fait, favorise l’invasion du tissu sain par les cellules tumorales. Une compréhension approfondie de la progression tumorale passe nécessairement par la caractérisation numérique de la géométrie de la FN fibrillaire.

Nous proposons dans cette étude une approche combinant

extraction de graphes et mesures statistiques pour caractériser la structure fibrillaire de la FN.

2 Données

Deux bases de données sont utilisées dans cette étude. La première est constituée d’images de cultures *in vitro*. Des fibroblastes d’embryons de souris ont été cultivés selon le protocole décrit dans [1]. Afin de mimer le phénotype de fibroblastes associés au cancer, les cellules ont été traitées avec du TGF- β 1, une cytokine prévalente dans le microenvironnement tumoral. Après sept jours, les cultures ont été décellularisées pour produire une MEC dérivée des cellules. La MEC a été imagée par microscopie confocale après marquage en immunofluorescence de la FN. Au total, 360 images (1000 \times 1000 pixels) sont utilisées dans cette étude, réparties de manière équilibrée entre deux classes.

La deuxième base de données est constituée d’images multispectrales d’immunofluorescence de tissus tumoraux de la tête et du cou provenant de l’essai clinique d’immunothérapie (Unicancer TOPNIVO-ORL09). Chaque image composite, de taille

1872 × 1404 pixels, est composée de sept canaux spectraux distincts, dont un correspondant à la FN.

L'utilisation des deux bases de données se justifie ainsi : les données *in vitro*, images homogènes et annotées, valident l'approche avant son application aux images du tissu tumoral, qui, quant à elles, sont caractérisées par une hétérogénéité et une complexité accrues.

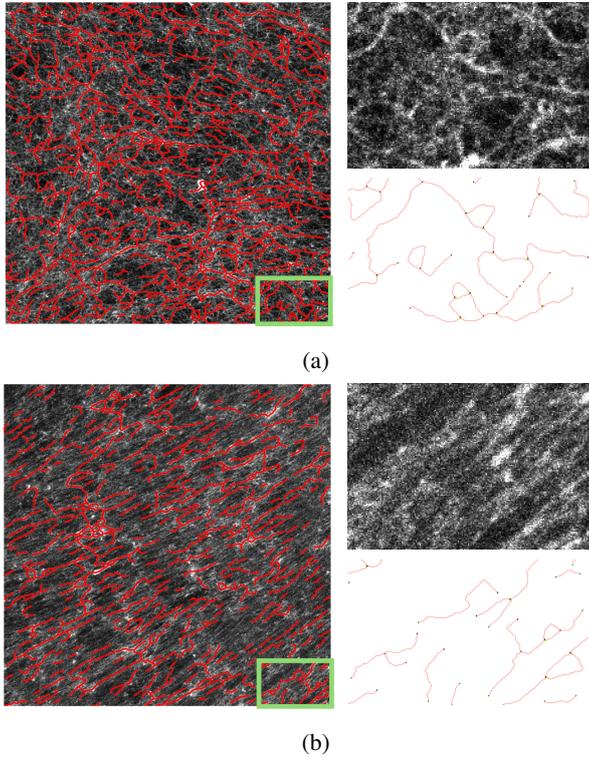


FIGURE 1 : Images de FN dans une MEC décellularisée produite par des fibroblastes (a) témoins ou (b) traités au TGF- β 1, mimant un phénotype tumoral, avec leur représentation graphique correspondante (en rouge).

3 Caractérisation des réseaux de FN

L'objectif final de cette étude est de proposer une batterie de mesures statistiques discriminantes permettant de caractériser numériquement le continuum organisationnel des fibres. À une extrémité du continuum, les fibres sont lâches et dépourvues d'orientation préférentielle, et à l'autre extrémité, elles sont fortement alignées et sous tension.

La chaîne de traitement proposée dans [2] est utilisée dans cette étude pour caractériser les réseaux fibrillaires. Le contraste de chaque image $u \in \mathbb{R}^{n \times m}$ est d'abord réhaussé avec une famille de noyaux de Gabor pour obtenir $\tilde{u}(x, y) := \max_{\theta, \sigma} (G_{\theta, \sigma} * u)(x, y)$. L'image réhaussée \tilde{u} est ensuite seuillée afin d'obtenir un squelette, qui peut être vu comme un résumé topologique des fibres sous-jacentes. Parallèlement, une carte d'orientation θ est également calculée correspondant à l'angle maximisant la réponse de Gabor en chaque pixel. Deux types d'élément structurel sont ensuite directement identifiés dans l'image squelettisée : les points terminaux, correspondant aux extrémités des fibres, et les points de bifurcation où les fibres se séparent. Ces éléments servent ensuite de nœuds au graphe topologique $G_u = (E_u, V_u)$, dont les arêtes représentent les connexions fibrillaires (Fig.1).

Enfin, une partition de l'image est obtenue en générant un diagramme de Voronoi à partir des nœuds $V_u = \{v_1^u, \dots, v_K^u\}$. Cette partition divise l'image en régions, chacune associée à un nœud du graphe, où chaque pixel appartient à la région du nœud le plus proche. De ce fait, chaque cellule de Voronoi peut être vue comme une fenêtre adaptative permettant une caractérisation multi-échelle, localement via les propriétés géométriques de chaque cellule, et globalement par la topologie du graphe sous-jacent. La cellule de Voronoi du site v_k^u , définie comme l'ensemble des points plus proches de v_k^u que de tout autre site, est notée \mathcal{V}_k .

Une caractéristique clé distinguant les classes est l'alignement des fibres (Fig. 1). Pour quantifier cet alignement, nous définissons un score η_k pour chaque cellule de Voronoi \mathcal{V}_k de l'image :

$$\eta_k = \left\| \frac{\sum_{x, y \in \mathcal{V}_k} \tilde{u}(x, y) \mathbf{s}(x, y)}{\sum_{x, y \in \mathcal{V}_k} |\tilde{u}(x, y)|} \right\|_2$$

où $\mathbf{s}(x, y) = (\cos 2\theta(x, y), \sin 2\theta(x, y))'$ est le vecteur unitaire associé à l'orientation locale $\theta(x, y)$, $\tilde{u}(x, y)$ est l'intensité du pixel (x, y) dans l'image réhaussée \tilde{u} , et $\|\cdot\|_2$ désigne la norme euclidienne. Ce score satisfait $\eta_k \in [0, 1]$, avec $\eta_k = 1$ lorsque toutes les fibres sont parfaitement alignées, et $\eta_k \approx 0$ pour une distribution aléatoire des orientations.

La topologie des squelettes est caractérisée par les mesures suivantes : nombre de branches entre extrémités (End-to-End), nombre de branches entre jonctions et extrémités (Junction-to-Endpoint), nombre de branches entre jonctions (Junction-to-Junction) et la longueur des branches.

Chaque branche $\gamma: I \subset \mathbb{N} \rightarrow \mathbb{R}^2$ du squelette est caractérisée par sa tortuosité définie par $T(\gamma) := L(\gamma)/d(\gamma)$ où $L(\gamma) := \sum_n \|\gamma_{i_n} - \gamma_{i_{n-1}}\|_2$ est la longueur de la courbe, définie comme la somme des longueurs des arcs, et $d(\gamma) := \|\gamma_{i_n} - \gamma_{i_0}\|_2$ est la distance euclidienne entre les points de départ et d'arrivée de la courbe.

Les nombres de Betti, notés β_0 et β_1 , sont des mesures de la complexité topologique du graphe $G = (E, V)$. Le nombre de composantes connexes définit β_0 . Le nombre de cycles dans le graphe est donné par $\beta_1 := \beta_0 + |V| - |E|$.

Les mesures de centralité sont des mesures de l'importance d'un nœud donné dans un graphe G [6]. Nous considérons trois mesures dans cette étude :

1. La centralité d'intermédiarité mesure la fréquence à laquelle un nœud apparaît sur les plus courts chemins entre une paire de nœuds donnée dans le graphe. Elle est définie par :

$$c_B(u) = \sum_{s \neq t} \frac{\sigma(s, t|u)}{\sigma(s, t)} \quad (1)$$

où $\sigma(s, t)$ est le nombre total de plus courts chemins allant du nœud s au nœud t , et $\sigma(s, t|u)$ est le nombre de ces chemins qui passent par u .

2. La centralité de proximité est une mesure de la manière dont un nœud est connecté à tous les autres nœuds en fonction des plus courts chemins. Elle est définie par :

$$C_{WF}(u) = \frac{n^2}{|V| - 1} \left(\sum_{\substack{v \sim u \\ v \neq u}} d(u, v) \right)^{-1}$$

où $n = |\{v \in V : u \sim v, u \neq v\}|$ et $d(u, v)$ est la longueur du plus court chemin entre les sommets u et v .

- Un sommet $v \in V$ est un point d'articulation si sa suppression augmente le nombre de composantes connexes, c'est-à-dire si $G - v$ a plus de composantes connexes que G . La cardinalité de l'ensemble de ces sommets est utilisée pour mesurer la propension du graphe à se déconnecter par la suppression de sommets critiques.

4 Résultats

La discrimination entre configurations fibrillaires normales et tumorales repose sur l'hypothèse que leurs propriétés géométriques et topologiques sont fondamentalement distinctes. Les valeurs sont présentées dans le tableau 1 pour les 360 images.

Pour vérifier cette hypothèse quantitativement, nous évaluons la capacité discriminative des mesures statistiques proposées (Section 3), leur performance relative face à un réseau de neurones (Section 4.2), et leur applicabilité à la segmentation d'images de tissus tumoraux (Section 4.3).

TABLE 1 : Métriques des réseaux de fibres pour fibroblastes témoins vs tumoroïdes

Métrique	Contrôle	Tumoroïde
Nombre d'extrémités libres	126.20 ± 33.78	192.83 ± 36.75
Nombre jonctions-extrémités	224.44 ± 58.19	176.43 ± 45.43
Nombre jonctions-jonctions	379.61 ± 265.18	143.15 ± 78.38
Longueur moyenne des branches	32.32 ± 3.42	40.60 ± 4.18
Alignement moyen	0.331 ± 0.016	0.430 ± 0.061
Alignement médian	0.323 ± 0.020	0.438 ± 0.070
β_0	161.19 ± 39.84	23.76 ± 38.17
β_1	248.91 ± 161.68	105.37 ± 50.02
Intermédiarité moyenne	0.0026 ± 0.0042	0.0001 ± 0.0003
Proximité moyenne	0.0073 ± 0.0046	0.0033 ± 0.0011
Intermédiarité max.	0.0446 ± 0.0679	0.0031 ± 0.0067
Proximité max.	0.0201 ± 0.0077	0.0124 ± 0.0035
Tortuosité médiane	1.1169 ± 0.0124	1.1121 ± 0.0106

Remarque : Toutes les métriques présentent des différences significatives ($p < 0.001$).

4.1 Classification avec les caractéristiques géométriques

Plusieurs caractéristiques présentent des corrélations fortes, ce qui peut entraîner une redondance d'information. Pour chaque paire de caractéristiques fortement corrélées ($\rho \geq 0.9$), seule la caractéristique ayant la corrélation absolue la plus élevée avec la variable cible, c'est-à-dire le résultat binaire où $y_i = 1$ pour les assemblages de type tumeur et $y_i = 0$ pour la classe contrôle, est conservée. Afin de filtrer le sous-ensemble de caractéristiques résultant et de contrôler le surajustement, un modèle de régression logistique régularisé est ajusté. Cette filtration en deux étapes permet de supprimer les redondances triviales (lorsque $\rho \geq 0.9$) et d'optimiser ensuite finement la sélection via la fonction de régularisation. Soit x_i le vecteur contenant les p caractéristiques restantes correspondant à l'échantillon i , $i \in \{1, \dots, N\}$. Les paramètres optimaux

$(b_0^*, \mathbf{b}^*) \in \mathbb{R}^{p+1}$ sont donnés par la solution unique du problème d'optimisation suivant :

$$\arg \min_{b_0, \mathbf{b}} \left\{ \mathcal{R}(\mathbf{b}) - \sum_{i=1}^N \left[y_i (b_0 + \mathbf{b}' \mathbf{x}_i) - \log(1 + \exp(b_0 + \mathbf{b}' \mathbf{x}_i)) \right] \right\} \quad (2)$$

où $\mathcal{R}(\mathbf{b})$ est un terme de régularisation. Pour contrer la forte corrélation entre les différentes caractéristiques, nous utilisons un terme de régularisation Elastic-Net :

$$\mathcal{R}(\mathbf{b}) = \alpha (\lambda |\mathbf{b}|_1 + (1 - \lambda) |\mathbf{b}|_2^2) \quad (3)$$

où $\alpha \in \mathbb{R}_+^*$, $\lambda \in (0, 1)$ sont des paramètres de régularisation choisis par validation croisée.

Le coefficient $|b_i^*|$, pour $i \in \{1, \dots, p\}$, peut être interprété comme l'importance relative de la caractéristique correspondante, c'est-à-dire sa force d'influence sur la probabilité prédite de la classe, avec des valeurs absolues plus élevées ayant une influence plus forte sur la classe prédite (Fig. 2). Les caractéristiques ayant les valeurs d'importance les plus élevées sont évaluées en fonction de leur capacité à classer les images. Seules les trois statistiques les plus importantes au sens de la fonctionnelle 4.1 sont retenues dans ce qui suit.

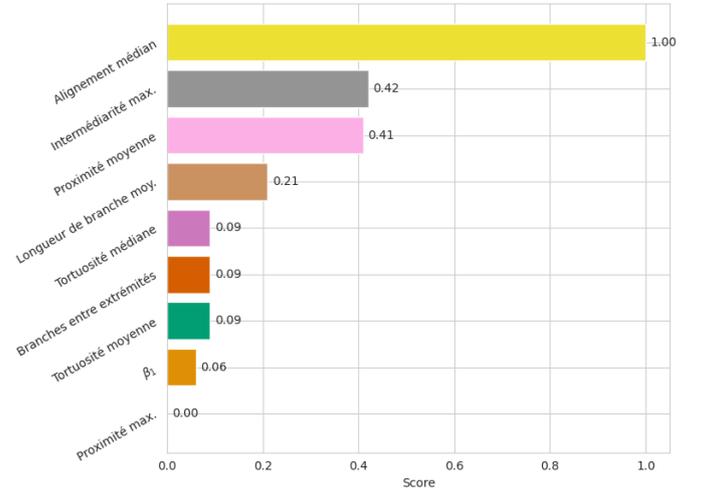


FIGURE 2 : Importance des mesures statistiques proposées

Chaque image est transformée, via la chaîne de traitement décrite précédemment, en une représentation vectorielle tridimensionnelle où chaque pixel encode l'alignement, la centralité d'intermédiarité et la centralité de proximité du site associé. Intuitivement, cela quantifie simultanément l'organisation géométrique des fibres et leur connectivité topologique en chaque pixel. Cette représentation (3D par pixel) est ensuite projetée dans un espace de 200 dimensions via Isomap, une technique de réduction de dimensionnalité non linéaire. Les caractéristiques réduites sont utilisées pour entraîner et évaluer plusieurs classificateurs. Le F1-score vaut 90%.

4.2 Classification avec le réseau EfficientNet

Nous évaluons la performance des caractéristiques proposées en les comparant à celles extraites par EfficientNet [5] dans un cadre de classification. Les résultats de cette comparaison sont présentés dans le tableau 2.

Le jeu de données ($n = 360$) est divisé en trois sous-ensembles : entraînement (70%), validation (15%) et test (15%). Des techniques d’augmentation des données, comprenant des rotations, des translations, des symétries axiales et des zooms, sont appliquées au sous-ensemble d’apprentissage. Le modèle est optimisé à l’aide de l’algorithme Adam, avec un taux d’apprentissage initial fixé à 10^{-4} et l’entropie croisée catégorielle comme fonction de perte.

Nous utilisons une approche d’apprentissage par transfert en prenant EfficientNetV2S comme modèle de base, pré-entraîné sur ImageNet, sans sa couche supérieure. Le modèle est enrichi d’une couche de Global Average Pooling 2D, suivie d’une couche dense (1024 unités, activation ReLU, régularisation L2), d’une normalisation par lots (Batch Normalisation), d’une couche Dropout (0.4) et d’une couche de sortie softmax. Le modèle de base est gelé pendant la phase initiale d’entraînement et affiné dans les étapes ultérieures. La performance est évaluée en utilisant le F1-score sur l’ensemble de test. Celui-ci vaut 85.71%.

TABLE 2 : Comparaison des F1-Scores

Method	F1-score
Statistiques proposées avec Forêt aléatoire	90.00%
EfficientNet	85.71%

4.3 Application aux images de tissus tumoraux

Compte tenu de la pertinence des statistiques proposées, comme en témoignent les tableaux 1 et 2, nous segmentons les images de tissu tumoral en utilisant les mesures suivantes : le score d’alignement, la centralité d’intermédiarité et la centralité de proximité. Concrètement, chaque cellule de Voronoi est associée à une classe prédite via un algorithme de K -means appliqué aux caractéristiques du site correspondant (Fig.3). Cette approche transpose aux images issues d’un essai clinique (plus hétérogènes et non annotées) les descripteurs discriminants identifiés et validés sur des images *in vitro* annotées (cf. Figure1), où la distinction entre FN contrôle et FN tumorale est connue. Elle illustre ainsi le potentiel de généralisation de notre méthode dans un contexte biomédical réel.

5 Conclusion

La fibronectine (FN) présente une organisation hétérogène, rendant difficile la caractérisation individuelle des fibres. Nous proposons une approche alternative partitionnant l’image en régions homogènes basées sur la topologie fibrillaire, permettant une caractérisation locale du comportement collectif des fibres.

Les mesures les plus discriminantes de la MEC *in vitro* entre conditions normales et tumorales incluent l’indice d’alignement, les centralités (intermédiarité, proximité) et la longueur des branches, en accord avec la littérature [3]. Notre méthode, fondée sur des caractéristiques explicites, surpasse une approche state-of-the-art (F1-Score = 90%), tout en offrant une interprétabilité biologique cruciale.

Validée sur des images cliniques malgré leur complexité accrue, cette approche ouvre la voie à des marqueurs pronostiques pour la stratification des patients et la prédiction de résistance à l’immunothérapie.

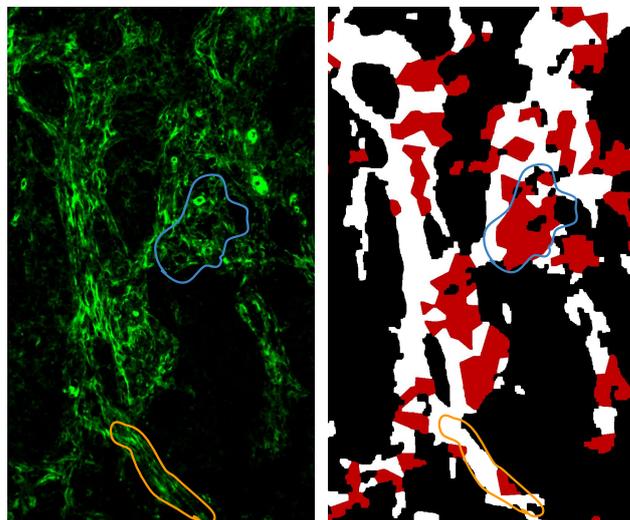


FIGURE 3 : Segmentation de la FN stromale dans la MEC tumorale. L’image de gauche montre la FN en vert, avec deux régions annotées : en orange, une zone à fort alignement et connectivité ; en bleu, une zone désorganisée. L’image de droite montre la classification automatique des cellules de Voronoi : les régions blanches correspondent aux structures alignées, les rouges aux structures désorganisées. La zone bleue reflète un profil tumoral.

Remerciements

Ce travail a été soutenu par le gouvernement français à travers les Investissements 3IA Côte d’Azur dans le cadre de la subvention ANR-23-IACL-001 de l’Agence Nationale de la Recherche (ANR) et par le Centre National de la Recherche Scientifique (CNRS) à travers les programmes interdisciplinaires de la MITI. Les auteurs remercient également toutes les personnes ayant contribué aux essais cliniques d’Unicancer TOPNIVO et CHECKUP ainsi qu’à la génération des images multispectrales.

Références

- [1] Anca-Ioana Grapa et AL. : A spatial statistical framework for the parametric study of fiber networks : Application to fibronectin deposition by normal and activated fibroblasts. *Biological Imaging*, 3:e25, 2023.
- [2] Faisal Jayousi et AL. : Detection and characterisation of fibronectin structures in the tumour extracellular matrix. *In EUSIPCO 2024, Lyon, France*, pages 1656–1660, 2024.
- [3] Paolo P. et al. PROVENZANO : Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Medicine*, 4(1):38, Dec 2006.
- [4] Purva SINGH, Cara CARRAHER et Jean E. SCHWARZBAUER : Assembly of fibronectin extracellular matrix. *Annu. Rev. Cell Dev. Biol.*, 26:397–419, 2010.
- [5] Mingxing TAN et Quoc V. LE : Efficientnet : Rethinking model scaling for convolutional neural networks. *ArXiv*, abs/1905.11946, 2019.
- [6] Stanley WASSERMAN et Katherine FAUST : *Social Network Analysis : Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.