



HAL
open science

Analyse du comportement collectif à l'aide de graphes, application à l'évaluation de l'indépendance des trajectoires de navires de pêche

Jérôme Roux

► **To cite this version:**

Jérôme Roux. Analyse du comportement collectif à l'aide de graphes, application à l'évaluation de l'indépendance des trajectoires de navires de pêche. Sciences agricoles. Nantes Université, 2023. Français. NNT : 2023NANU4045 . tel-04595092

HAL Id: tel-04595092

<https://theses.hal.science/tel-04595092>

Submitted on 30 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 642

Ecole doctorale Végétal, Animal, Aliment, Mer, Environnement

Spécialité : Statistique, Modélisation, Intelligence artificielle en sciences du vivant et géosciences

Par

Jérôme ROUX

Analyse du comportement collectif à l'aide de graphes, application à l'évaluation de l'indépendance des trajectoires de navires de pêche.

Thèse présentée et soutenue à Nantes, Ifremer - Centre Atlantique, le 07 Juillet 2023

Unité de recherche : Ifremer, UMR DECOD, Dynamique et durabilité des écosystèmes : de la source à l'océan

Rapporteurs avant soutenance :

Catherine MATIAS Directrice de recherche, CNRS-Sorbone Université, Paris
Ronan FABLET Professeur de l'Ecole des Mines - IMT, IMT Atlantique, Brest

Composition du Jury :

Président :	Sophie DONNET	Directrice de recherche, INRAE - MIA, Paris-Saclay
Examineurs :	David NERINI	Maître de conférences des universités, OSU Pythéas, Marseille
	Lionel RICHEFORT	Maître de conférences des universités, IAE Nantes Université, Nantes
Dir. de thèse :	Stéphanie MAHEVAS	Cadre de recherche, Ifremer, Sète
Co-enc. de thèse :	Julien LEBRANCHU	Ingénieur de recherche, IRD, Sète
Co-enc. de thèse :	Nicolas BEZ	Directeur de recherche, IRD, Sète

Invité(s) :

Julien LEBRANCHU Ingénieur de recherche, IRD, Sète

AVANT-PROPOS

La vie oscille, comme un pendule, de droite à gauche, de la souffrance à l'ennui.
(A.Schoppenhauer[1])

La réalisation d'une thèse est un exercice de longue haleine souvent comparée à un marathon. Mais de quel marathon parle-t-on ? S'agit-il du marathon de Paris, 42 km de plats où les avenues rectilignes s'encastrent entre des façades haussmanniennes ternes et sans fantaisies et où les seules bouffées d'oxygène s'incarnent par de brèves incursions vertes dans le bois de Vincennes ou de Boulogne ? Ou s'agit-il du marathon de la Grande Muraille de Chine, 42km d'oscillations, 1700m de dénivelé positif répartis en plus de 3700 marches de pierre montantes ou descendantes, interrogeant l'ampleur du travail de forçat de Sisyphe[2] ? Qu'il soit plat ou cabossé chaque doctorant a son marathon...

On entend aussi souvent dire, que tout comme un marathonien concentré sur sa course, entrant dans sa bulle, les doctorants finissent par se couper du monde, obnubilés par leur thèse du lever au coucher, la réalité glisse sur eux et un matin, la bulle éclate, trois (ou quatre ?) années se sont écoulées... Dans mon cas, la bulle ne s'est jamais formée, ou bien trop fine elle a vite éclatée, ne résistant pas au tranchant de la réalité ou plus précisément, à son piquant... Il apparaît d'une ironie sans commune mesure que ce "confinement mental" ait été interdit, ou ait pris fin prématurément, par un confinement physique cette fois-ci.

Passée la pandémie, c'est la chaleur de la dérive climatique qui glace le sang. Incendie en Australie, dômes de chaleur au Canada, et que penser de cet été de sécheresse et de canicule ? 20 ans après le IV^e Sommet de la Terre[3] et son fameux "*Notre maison brûle et nous regardons ailleurs*", 25 ans après la signature du protocole de Kyoto[4], 50 ans après la publication du rapport du Club de Rome[5], et après tant de rapports du GIEC[6], il apparaît ici d'une ironie morbide et révoltante, que la "*fin de l'abondance, de l'insouciance et de l'évidence*" soit incarnée par la fermeture d'un simple robinet de gaz imposant l'appel à la sobriété.

L'humanité ne semble être capable de réagir que sous la contrainte et nous allons semble-t-il devoir encore beaucoup réagir. Le 21^e siècle s'annonce être le siècle de tous les "pics" ("*Peak Everything*") [7], celui du pétrole est déjà passé en 2006 (du moins pour

sa forme conventionnelle)[8], celui du gaz semble également être passé ou bien est en passe de l'être, le pic du phosphore essentiel à l'agriculture se profile pour 2030, celui de la production de ciment limitée par l'extraction de sable serait passé en 2018, sans compter les pics de tous les métaux indispensables à nos technologies, lithium, argent, cuivre, zinc, etc. Une question reste en suspens, le pic du CO_2 aura-t-il lieu au cours du 21^e siècle? Certainement que oui, mais une question demeure, quelle hauteur aura-t-il atteint lorsqu'il se produira?

Le début de ce siècle dont un quart ne n'est pas encore écoulé résonne étrangement avec *l'introduction au siècle des menaces*[9]. Tandis que nous regardons notre sol craquelé par la sécheresse, d'autres regardent vers les étoiles ou plutôt vers Mars. Quel étrange désir de vouloir quitter notre si belle planète pour une autre, l'herbe semble toujours plus verte ailleurs, mais savent-ils seulement qu'il n'y a pas d'herbe sur Mars? A moins que leurs regards remplis d'envie ne camouflent une peur profonde de l'avenir, il paraît que lorsque la brèche ne peut plus être colmatée, les rats quittent le navire.

Le point de chute de l'humanité est tout tracé, après deux siècles d'une succulente sensation d'apesanteur sous tendue par des croissances exponentielles, c'est un retour à l'état d'équilibre qui s'impose pour nous. Seul le chemin qui nous y mènera reste à définir, déciderons nous de choisir notre chemin ou laisserons nous la "Nature" s'en charger? Guerre, famine et maladie, redoutables briseuses d'exponentielles en sont ses armes et nous, quelles sont les nôtres? Des avions volants à l'huile de friture, des voitures électriques de deux tonnes, un réseau 5G mondial ou des panneaux solaires produits au charbon? Soyons sérieux...

Dans ce contexte ne peut-on pas interroger la pertinence d'une thèse en halieutique? N'apparaît-il pas dérisoire d'interroger les impacts et d'élaborer des stratégies de gestion d'un système d'exploitation non durable? Non durable non pas au regard de la ressource exploitée, mais au regard de la dépendance des exploitants aux énergies fossiles, elles-mêmes non durables par "essence". Peut-être faudrait-il songer à ajouter les chalutiers à la liste des espèces en voie de disparition... Il semble en effet peu probable que l'exploitation halieutique apparaisse aussi intense qu'elle ne l'est actuellement, avec des flottilles de chalutiers électriques...

Le lecteur pardonnera, je l'espère, le ton peut-être non conventionnel ou non académique de ce manuscrit, c'est le ton d'un jeune homme de 29 ans qui jeta son regard vers l'avenir, et à la manière de Poincaré face au chaos [10], en fut horrifié...

REMERCIEMENTS

Je tiens à remercier en premier lieu les deux financeurs de ce travail de thèse :

- La direction scientifique de l’Ifremer au travers du projet Tracflo.
- Et la région Pays de la Loire.

En second lieu, je remercie vivement toutes les personnes m’ayant accompagné et soutenu durant ces près de 4 années. Tout d’abord mes encadrants de thèse, je remercie chaleureusement Stéphanie Mahévas et Nicolas Bez pour leur gentillesse, leur soutien, leur écoute, leurs nombreuses remarques toujours constructives et pour m’avoir systématiquement aidé à revenir sur le droit chemin quand j’avais tendance à m’égarer, ainsi que Julien Lebranchu pour son rôle particulier dans mon encadrement.

Je souhaite remercier ensuite les membres de mon comité de suivi de thèse, Sophie Bertrand, Hilaire Drouineau et Paul Rochet pour leur curiosité, leur disponibilité et leur regard critique sur mon travail ayant conduit à des échanges riches d’enseignements.

Enfin, je remercie particulièrement trois personnes m’ayant accompagné à différents moments de ma thèse, à nouveau Paul Rochet qui m’a accompagné pendant presque toute sa durée, je le remercie ici pour sa grande bienveillance et son expertise de l’analyse des graphes, Rocío Joo qui a été également très présente, je la remercie pour son écoute et son expertise de l’analyse des trajectoires et Amédée Roy pour son aide sur la fin de ma thèse et son expertise des modèles de GAN. Je vous remercie sincèrement pour vos échanges et pour votre disponibilité ayant contribué à rendre mon travail meilleur.

Je remercie également l’Ifremer et particulièrement l’ensemble de l’unité EMH pour son accueil chaleureux. La liste serait trop longue pour tous les mentionner, alors qu’ils soient stagiaires ou chercheurs, de l’EMH ou non, je remercie particulièrement tous ceux qui m’ont soutenu et qui ont quotidiennement ensoleillé mon ciel pluvieux.

Finalement, je ne peux que remercier du fond du coeur ma famille et en particulier mes parents sans qui toutes ces années d’études n’auraient pas été rendues possible.

TABLE DES FIGURES

1.1	Le planisphère de Cantino	2
1.2	Les routes commerciales dans l'océan mondial	4
1.3	Évolution de la production halieutique et aquacole mondiale	6
1.4	Évolution du niveau d'exploitation des stocks de ressources marines	7
1.5	Exemple d'histogrammes des vitesses de navires de pêche	12
1.6	Segmentation d'une marée par la vitesse	13
1.7	Illustration de dyades	14
1.9	Les différents types de graphes	17
1.10	Le graphe du Karate club de Zachary	18
1.11	Le graphe des sept ponts de Königsberg d'Euler	20
1.12	Les graphes du monastère de Sampson	22
1.13	Série temporelle des publications scientifiques sur les modèles de graphes	24
1.14	Exemples de réalisations des modèles Erdős-Rényi et Gilbert	25
1.15	Les phases de transition du modèle d'Erdős-Rényi	27
1.16	La distribution binomiale des degrés d'un graphe d'Erdős-Rényi	28
1.17	Évolution du diamètre des graphes du modèle d'Erdős-Rényi	30
1.18	Illustration de la réalisation du modèle de Watts-Strogatz	32
1.19	Évolution du diamètre et de la transivité d'un graphe de Watts-Strogatz	33
1.20	Loi de puissance " <i>scale-free</i> "	34
1.21	Illustration de la réalisation du modèle de Barabási-Albert	35
1.22	Résumé graphique	38
2.1	Carte de la Manche	40
2.2	Les chalutiers en boeufs (PTM)	42
2.3	Les chalutiers de fond (OTB)	42
2.4	La double discrétisation des données VMS	44
2.5	Histogrammes des durées des marées PTM et OTB	46
2.6	Illustration d'un ensemble de trajectoires PTM ou OTB	47
2.7	Les trois métriques descriptives des dyades	49
2.8	Histogrammes des pas de temps d'échantillonnage VMS	52

2.9	Asynchronie des fréquences d'échantillonnage VMS	52
2.10	Interpolation linéaire d'une trajectoire VMS	54
2.11	Exemple de trajectoires incomplètes	56
2.12	Distributions des métriques descriptives des dyades PTM et OTB	59
2.13	Illustration d'un modèle de mélange gaussien GMM	60
2.14	Distributions des métriques des dyades par groupe de cohésion	63
2.15	ACP des dyades PTM et OTB	64
2.16	Illustration des graphes des dyades PTM	65
2.17	Graphes des dyades entres les marées ou les navires	66
2.18	Composante temporelle des graphes de dyades	67
2.19	Composante temporelle des graphes de dyades	68
3.1	Graphes d'illustration des méthodes de comparaisons de graphes	74
3.2	Vecteur signature NetSimile	78
3.3	Effet de la densité sur les caractéristiques des distributions des métriques topologiques	79
3.4	Spectres d'amplitudes des valeurs propres	80
3.5	Graphlets d'ordre 2 à 5	83
3.6	Exemple de décomposition des graphlets	83
3.7	Distributions des fréquences des graphlets	84
3.8	Exemple de décomposition des orbites	86
3.9	Exemple de matrice de corrélation des graphlets	88
3.10	Illustration de la segmentation hebdomadaire de la flottille OTB	92
3.11	Illustration de la segmentation hebdomadaire de la flottille PTM	93
3.13	Qualité de la discrimination entre les modèles (AUPR)	102
3.14	Évolution des 55 coefficients de corrélation des graphlets	103
3.15	Probabilité de distinction entre deux modèles de graphes aléatoires d'ordres et/ou de densités différents	105
3.16	ACP entre des graphes d'Erdős-Rényi et Fitness Scale-Free d'ordre $N =$ 100 et de densités différentes	106
3.17	Illustration d'un graphe empirique et de sa Matrice des Corrélations des Graphlets	108
3.18	Comparaison des coefficients de corrélation des graphlets d'un graphe em- pirique au modèle de graphes aléatoires	110
3.19	Domaine de séparabilité de la GCD_{11} et projection des graphes empiriques	114

3.20	p-value de la comparaison des graphes OTB au modèle d'Erdős-Rényi . . .	115
4.1	Illustration d'un réseau de neurones	120
4.2	Illustration de l'architecture d'un GAN	121
4.3	Exemple de trajectoires simulées par GAN	124
4.4	Caractéristiques globales d'un ensemble de trajectoires simulées	125
4.5	Représentation d'un ensemble de trajectoires simulées	126
4.6	Exemples de graphes de trajectoires indépendantes	127
4.7	Graphes de trajectoires indépendantes et domaine de séparabilité	129
4.8	Comparaison des graphes d'indépendance aux graphes d'Erdős-Rényi . . .	130
4.9	Comparaisons de la densité des graphes empiriques et des graphes de tra- jectoires indépendantes	132
4.10	Test statistique des graphes empiriques et des graphes de trajectoires in- dépendantes	133

TABLE DES MATIÈRES

1	Introduction	1
1.1	L'Homme et la mer	1
1.1.1	De l'isolement à l'hyperconnexion	1
1.1.2	De la solitude à la multitude	3
1.1.3	De la multiplication du trafic maritime à l'addition des impacts anthropiques	4
1.2	La recherche au service de la gestion des pêches	5
1.2.1	La pêche mondiale, un système sous tension ?	6
1.2.2	La gestion des pêches dans l'Union Européenne	8
1.2.3	La recherche scientifique au service de la gestion	9
1.3	Du suivi des pêcheurs à l'analyse de leur comportement	10
1.3.1	La donnée VMS	11
1.3.2	De la trajectoire au comportement individuel	11
1.3.3	Du comportement individuel au comportement deux à deux	13
1.3.4	Du comportement deux à deux au comportement N à N	14
1.4	Le graphe, un objet mathématique	16
1.4.1	Définition et formalisation d'un graphe	16
1.4.2	Des graphes empiriques aux modèles de graphes aléatoires :	23
1.5	Objectif et plan de thèse :	36
2	De la trajectoire au graphe	39
2.1	Cas d'étude et démarche	39
2.1.1	La Manche, un univers de contraintes	39
2.1.2	Flottes d'études	41
2.1.3	Hypothèse et démarche	43
2.2	De la trajectoire à la dyade	43
2.2.1	La donnée VMS, une succession de point dans le temps et l'espace	43
2.2.2	L'activité de pêche, une succession de marées	45
2.2.3	La cohésion deux à deux, la brique fondamentale	47

2.2.4	Synchronisation et filtrage de la donnée VMS	51
2.3	Intensité et cohésion des relations dyadiques	57
2.3.1	Application aux flottilles d'études	58
2.3.2	Classification des relations dyadiques	59
2.4	Des relations 2 à 2 aux relations n à n	64
2.4.1	Le graphe de flottille, un objet modulable	64
2.4.2	Le graphe de flottille, un objet temporel	66
2.4.3	Le graphe de flottille, un objet spatial	68
2.5	Conclusion du chapitre	69
3	De l'indépendance des trajectoires à l'indépendance des relations dyadiques	71
3.1	Système d'hypothèse et démarche :	71
3.2	La similarité entre graphes	72
3.2.1	Quelle similarité?	72
3.2.2	L'isomorphisme, une similarité binaire	74
3.2.3	Les "Graph Edit Distances" (GED)	75
3.2.4	Les "features-based distances"	76
3.2.5	Les distances spectrales	79
3.2.6	Les "graphlet-based distances"	82
3.3	La comparaison de graphes, un problème de taille	90
3.3.1	Vers une réduction spatiale et temporelle des graphes de flottille	90
3.3.2	Vers la comparaison des petits graphes	95
3.3.3	Application, l'indépendance des graphes de flottilles	113
3.4	Conclusion du chapitre	115
4	De l'indépendance des relations à l'indépendance des trajectoires	117
4.1	Système d'hypothèse et démarche :	117
4.2	Vers des graphes de trajectoires indépendantes	118
4.2.1	Le modèle des réseaux antagonistes génératifs (GAN)	119
4.2.2	Des trajectoires indépendantes au graphe	122
4.3	Graphes de trajectoires indépendantes <i>versus</i> graphes d'arêtes indépendantes (Erdős-Rényi)	128
4.3.1	Graphes de trajectoires indépendantes et domaine de séparabilité	128
4.3.2	Comparaison aux graphes d'indépendance	130

4.4	Graphes de trajectoires indépendantes <i>versus</i> graphes de flottilles	131
4.4.1	Comparaison du nombre de rencontres	131
4.4.2	Comparaison des graphes empiriques aux graphes des trajectoires indépendantes	132
4.5	Conclusion du chapitre	134
	Discussion générale	135
	Informations supplémentaires	141
	Bibliography	145

INTRODUCTION

1.1 L'Homme et la mer

L'appel du large

Un matin nous partons, le cerveau plein de flamme,
Le coeur gros de rancune et de désirs amers,
Et nous allons, suivant le rythme de la lame,
Berçant notre infini sur le fini des mers,

Mais les vrais voyageurs sont ceux-là seuls qui partent
Pour partir, coeurs légers, semblables aux ballons,
De leur fatalité jamais ils ne s'écartent,
Et sans savoir pourquoi, disent toujours : Allons !

Amer savoir, celui qu'on tire du voyage !
Le monde, monotone et petit, aujourd'hui,
Hier, demain, toujours, nous fait voir notre image :
Une oasis d'horreur dans un désert d'ennui !

(C.Baudelaire [11])

1.1.1 De l'isolement à l'hyperconnexion

Qu'il est loin le temps des grandes traversées, des grandes expéditions maritimes où la sortie du port était synonyme d'entrée dans un monde inconnu. Qu'il est loin ce temps où pour celles et ceux restés à terre, le voyage s'arrêtait à la ligne d'horizon, passé cette ligne, c'est l'attente d'un retour incertain qui commence. Qu'il est loin ce temps où le chemin parcouru par ces marins partis dans l'autre monde n'était esquissé que dans l'imaginaire ou le monde onirique de celles et ceux qui s'endormaient à terre.

Ulysse traversant la Méditerranée de part en part [12], Christophe Colomb traversant

l'Atlantique [13], Louis-Antoine de Bougainville premier français à réaliser une circumnavigation [14], qu'ils soient fictifs ou réels, ces navigateurs savaient-ils seulement ce qu'ils faisaient ? Certains plus que d'autres probablement, mais une chose demeure certaine, une fois la dernière lueur du dernier phare disparue, tous ces navigateurs, bons ou mauvais, étaient seuls... désespérément seuls dans *une oasis d'horreur, dans un désert d'ennui!* [11].

Et puis, à compter du 15^{ème} siècle, avec l'essor des grandes découvertes [15], les bords du monde se sont joints, les lignes des côtes se sont affinées et les distances se sont raccourcies. L'âge des découvertes était également l'âge des cartographes (Fig 1.1) où les mystères entourant les océans, du moins ceux de la surface, s'amenuisaient à chaque réédition d'une nouvelle carte marine. Dans le même temps, les instruments de navigation se sont perfectionnés, l'astrolabe a laissé sa place au sextant et la boussole au compas de marine. Cartes marines et instruments de navigation plus précis ont permis aux navigateurs d'établir des routes plus courtes et plus sûres mais n'ont en rien résolu l'isolement du grand large.

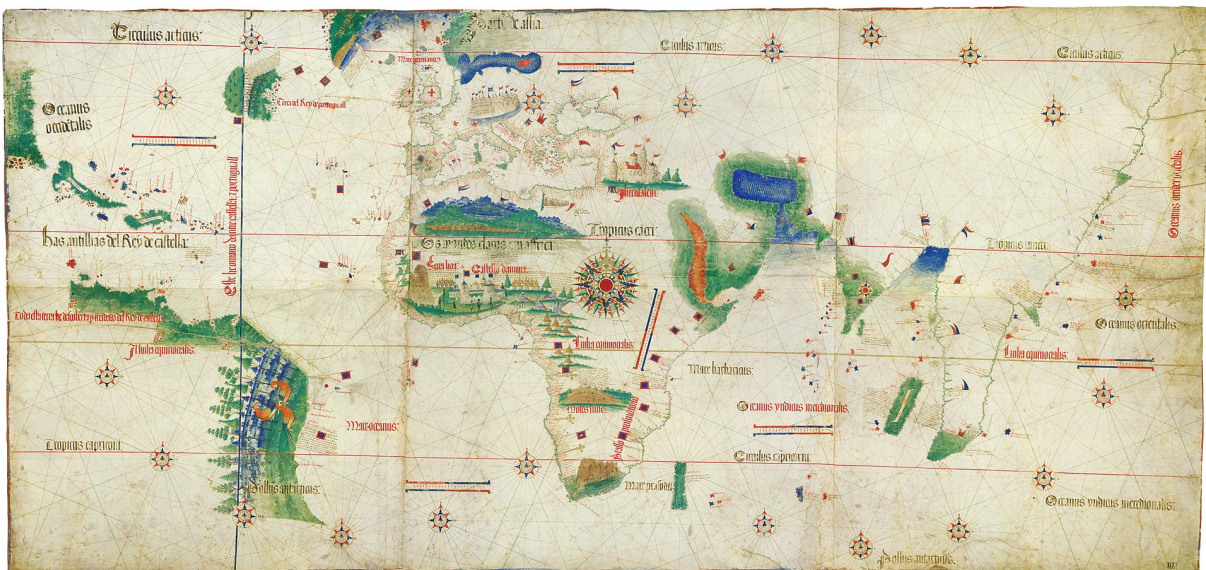


FIGURE 1.1 – Le planisphère de Cantino (Alberto Cantino, 1502).

Il faut attendre la seconde moitié du 20^{ème} siècle et l'apparition des premiers systèmes électroniques embarqués pour combler le silence de la haute mer et rompre définitivement l'isolement. Ainsi, l'apparition des outils de radionavigation, tels que les radiophares [16] et autres GPS ("*Global Positioning System*") [17], ont permis une géolocalisation d'une

grande précision en temps réel, les navigateurs ne sont désormais plus guidés par des constellations d'étoiles mais par des constellations de satellites. L'apparition des outils d'exploration, tels que les radars embarqués, les sondeurs bathymétriques et autres stations météorologiques, ont permis d'appréhender l'environnement à la fois sur et sous la surface de l'océan. Les masses d'eau, jusqu'alors opaques à l'œil humain, sont traduites en images sous la forme d'échos et de lignes bathymétriques. Et enfin, l'apparition des outils de communication longue distance : exit les lampes Aldis [18] et l'alphabet Morse, place aux communications VHF ("*Very High Frequency*") et satellitaires.

Tous ces outils électroniques d'aide à la navigation sont de nos jours monnaie courante à l'échelle de la flotte mondiale, du plus imposant des supertankers au plus petit des plaisanciers (du moins ceux des pays développés). Ces outils permettent non seulement de rendre compte du milieu environnant, mais également d'être en prise directe avec ceux qui s'y trouvent, favorisant ainsi les interactions et le dialogue entre les usagers de la mer. Ainsi équipé, plus aucun navire n'est véritablement seul où qu'il se trouve sur le globe.

1.1.2 De la solitude à la multitude

En plus d'être marquée par l'électronisation des navires à l'échelle mondiale, la deuxième moitié du 20^{ème} siècle est également marquée par l'incroyable densification de son trafic maritime. En premier lieu, l'essor de la marine marchande a permis l'interconnexion généralisée du monde (Fig 1.2), faisant des mers et des océans le support des flux soutenant la mondialisation [19]. En 2020, près de 90% du commerce international transitait par voies maritimes [20], c'est ainsi pas moins de 120 000 navires marchands spécialisés [21] (porte-conteneurs, vraquiers, pétroliers, méthaniers, etc.) qui ont assuré le transport de 10 milliards de tonnes de marchandises contre 550 millions de tonnes en 1950.

En second lieu, à la faveur des Trente Glorieuses [22], la navigation de plaisance a connu la même évolution. En France, par exemple, le nombre d'immatriculation de navires de plaisance est passé de quelques milliers en 1950 à plus de 800 000 en 2000 [23]. La mer n'est désormais plus un espace à conquérir réservé aux professionnels. C'est un espace de loisir à la portée de tous où ses plus beaux bijoux, ses plus belles baies, ses plus belles criques, souvent les plus reculés et secrets, sont exposés aux yeux du monde sur les réseaux sociaux.

Enfin, entre la flotte des navires marchands et celle de plaisance, se faufile une dernière flotte, plus discrète, elle se fond dans le paysage et se repère au troupeau de mouettes

qui l'accompagne : la flotte des navires de pêche. Pour répondre au besoin croissant de l'humanité en produits de la mer, la flotte mondiale des navires de la pêche industrielle ou artisanale a plus que doublé entre 1950 et 2015 passant, de 1,7 à 3,7 millions de navires [24].

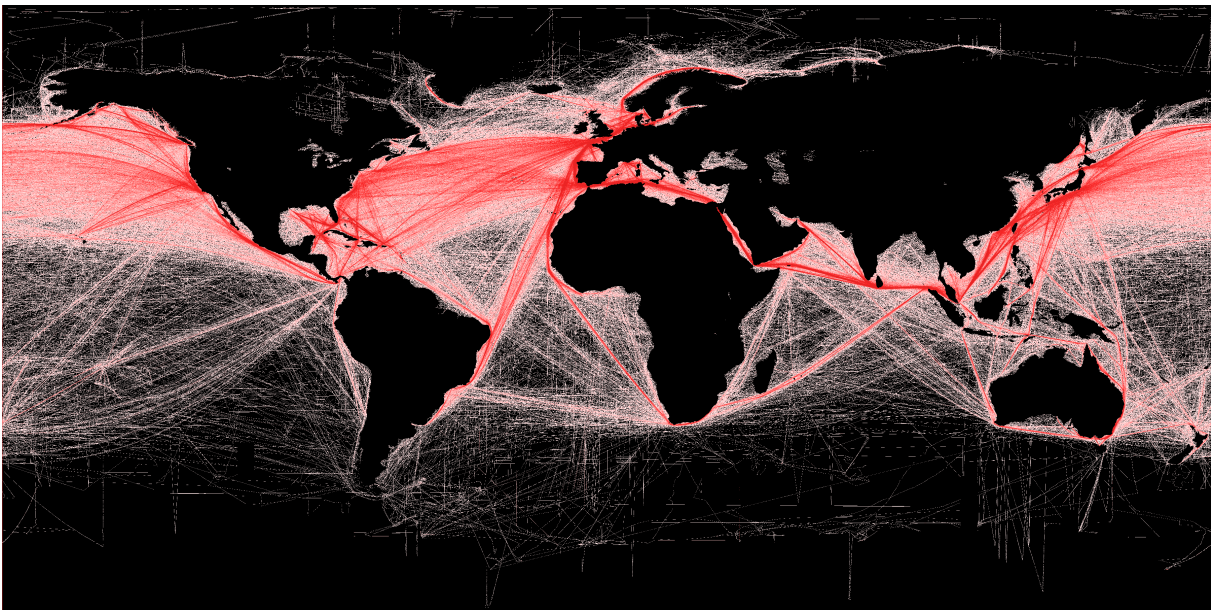


FIGURE 1.2 – Carte des routes commerciales dans l'océan mondial (B.S. Halpern, 2012).

Qui aurait pu prédire que l'immensité de l'océan serait un jour colonisée de la sorte par des bateaux qui se suivent, se doublent ou se croisent, rendant indispensable l'utilisation de radars anti-collision comme le système AIS ("*Automatic Identification System*" [25]) ?

1.1.3 De la multiplication du trafic maritime à l'addition des impacts anthropiques

A l'heure où les questions de soutenabilité et de durabilité des activités humaines sont sur le bord de toutes les lèvres, une telle transformation et densification du trafic maritime mondial n'est pas passée inaperçue et a fait émerger de nombreux questionnements. D'une part, des questionnements centrés sur la soutenabilité économique d'une telle organisation, interrogeant la résilience [26, 27] du réseau de transport maritime, véritable système sanguin de la mondialisation, dont l'épisode récent du blocage du canal de Suez [28], tel un caillot dans une artère, tend à démontrer sa vulnérabilité. D'autre part, des questionnements centrés sur la soutenabilité environnementale d'un tel trafic, interrogeant

ses impacts sur les écosystèmes marins [29], aussi bien ceux directement issus de la navigation, tels que la pollution sonore [30] ou les effets des ancrages en milieu naturel [31], que ceux issus des infrastructures nécessaires à cette navigation, tel que les effets de la multiplication ou de l'agrandissement des édifices portuaires [32]. Enfin, les impacts du trafic maritime s'inscrivent également dans des problématiques environnementales plus larges comme sa contribution aux émissions de gaz à effet de serre [33].

Parmi les trois grandes activités maritimes mentionnées précédemment (marchande, pêche ou plaisance), l'activité de pêche occupe une place tout à fait particulière dans le bilan des impacts anthropiques sur les écosystèmes marins. En effet, du point de vue des navires marchands ou de plaisance, la principale source d'impact est incarnée par leur seule activité de navigation (hors coût de construction ou événements exceptionnels, marées noires, déballastages ou même naufrages), faisant des mers et des océans un simple support du déplacement et rendant leurs impacts proportionnels à leurs "quantités de mouvement". Or, bien qu'elle se retrouve également chez les navires de pêche, l'activité de pure navigation et les impacts qui en découlent deviennent secondaires face à l'activité principale de ces navires, l'exploitation directe des ressources marines, la pêche.

1.2 La recherche au service de la gestion des pêches

La pêche est une pratique d'exploitation des ressources naturelles visant à prélever une ressource vivante dans son milieu naturel à l'aide d'engins de pêche (de capture) d'une grande diversité (casier, ligne, drague, chalut, senne...) dépendant des espèces recherchées. Sous tendue par la vision d'une ressource inépuisable¹, la pratique de la pêche s'est accrue et intensifiée.

Toutefois, depuis la fin de la seconde guerre mondiale, les expressions d'inquiétudes face au risque de surexploitation de certaines zones de pêches se sont multipliées comme en témoigne, par exemple, la prise de parole d'Edward Evans, député britannique, devant la Chambre des communes du Royaume-Uni en 1947 : "*Je souhaite que l'Assemblée se penche sur le danger qui semble se profiler en ce qui concerne la surpêche en mer du Nord et, bien que je souhaite limiter la portée de mes remarques à cette zone, les considérations que j'espère présenter s'appliquent non seulement à la mer du Nord mais à toutes les eaux*

1. "*Les pêcheries de morues, les pêcheries de harengs, les pêcheries de maquereaux, et probablement les grandes pêcheries maritimes sont inépuisables, ce qui veut dire que rien que nous faisons affecte sérieusement le nombre de poissons* , Huxley (1883)" [34]

situées à proximité de ces îles [...] Je ne prétends pas qu'il y aura un jour une famine de poisson, mais je crois que si une action résolue n'est pas entreprise rapidement, nous en souffrirons gravement, tant sur le plan de la qualité que sur celui de la variété.[...] La demande de poisson est si importante, non seulement dans ce pays, mais dans tous les pays européens voisins de la mer du Nord - en particulier en raison de la pénurie d'autres aliments - afin de compléter leurs réserves alimentaires, que l'intensité de la pêche dans les eaux proches dépasse la capacité d'absorption.[...] Après la guerre, tous les pays voisins ont intensifié leur pêche sans restriction et, en l'espace de trois ans environ, les pêcheries de la mer du Nord ont été réduites à néant. Il en a été de même pour d'autres parties de la côte" [35].

1.2.1 La pêche mondiale, un système sous tension ?

Le constat de l'intensification de la pêche dans la mer du Nord exprimé ci-dessus par le député Edward Evans et les inquiétudes d'un risque de surexploitation qui en découlent ne sont pas infondés à la seule mer du Nord mais peuvent être généralisés à l'échelle mondiale. En effet, d'après l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO), la production annuelle des produits de la mer (poissons, crustacés, mollusques) était de 20 millions de tonnes en 1950, pour atteindre en 2018, 180 millions de tonnes (Fig 1.3), soit un doublement de la production mondiale tous les 20 ans environ.

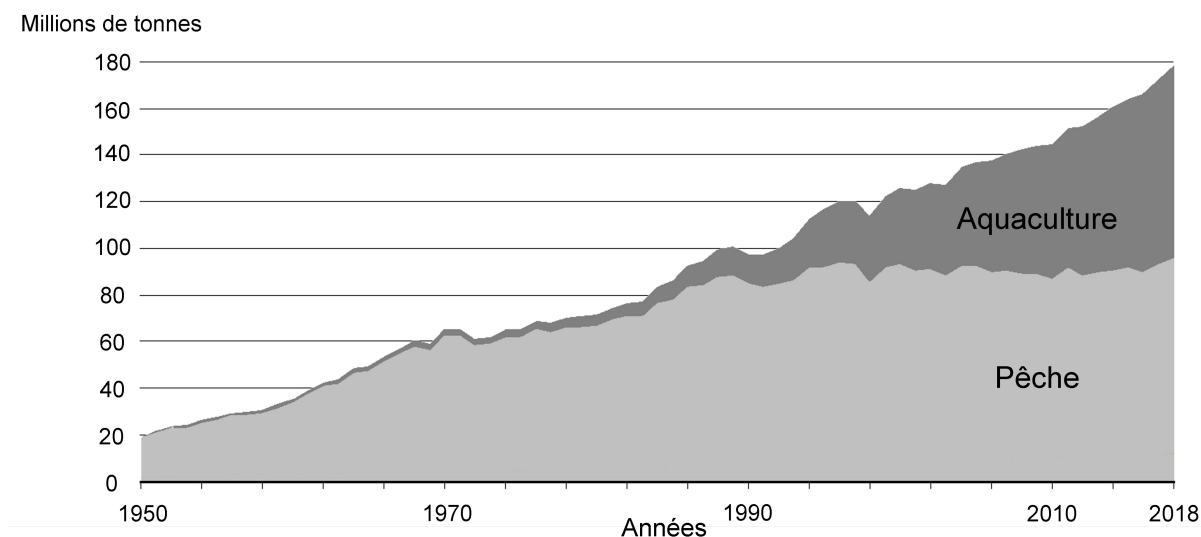


FIGURE 1.3 – Évolution de la production halieutique et aquacole mondiale (FAO [36]).

Cette croissance continue n'a pu s'opérer qu'avec l'apport de l'aquaculture passant d'une production anecdotique à la fin des années 1980 à une production en passe de devenir majoritaire en 2018 (46% de la production mondiale). Cette production aquacole n'a pas remplacé la production halieutique, mais est venue s'y superposer afin de répondre à la demande croissante de la population mondiale en produits de la mer. En effet, cette consommation est passée entre 1950 et 2018, de moins de 10kg à plus de 20kg par an et par habitant [37], avec dans le même temps, un quasi triplement de la population mondiale passant de 2,6 à 7,6 milliards d'individus.

L'évolution de l'aquaculture est intrinsèquement liée à celle des pêcheries et apparaît comme symptomatique d'une réponse au ralentissement de la production halieutique, signe que les limites de ce que peuvent offrir les écosystèmes marins ont été progressivement atteintes voire dépassées. L'effondrement de la pêcherie de morue de Terre-Neuve en 1992 [38] en est une des plus célèbres illustrations.

Cette érosion progressive des ressources halieutiques s'est traduite par une dégradation de l'état des stocks à l'échelle mondiale. Ainsi, toujours d'après la FAO, en 1974, 90% des stocks mondiaux de poissons se situaient à un niveau d'exploitation biologiquement durable, tandis qu'en 2017 ils étaient moins de 66% [37] (Fig 1.4). En 1974, 10% des stocks mondiaux étaient exploités à des niveaux biologiquement non durable, tandis qu'en 2017, ils étaient plus de 34%.

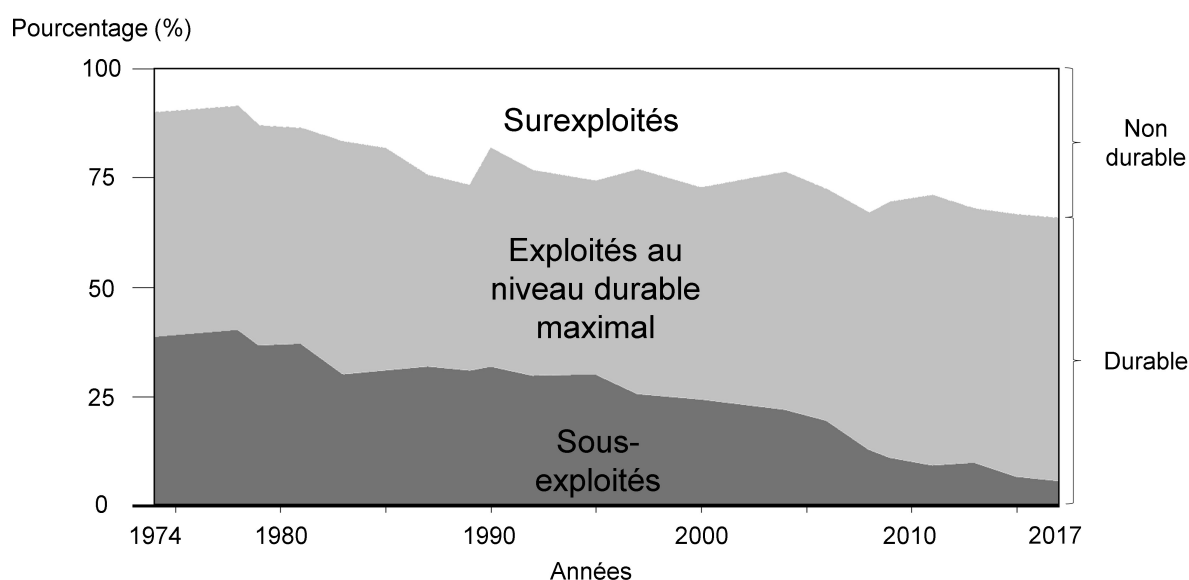


FIGURE 1.4 – Évolution de l'exploitation des stocks halieutiques mondiaux (FAO [37]).

Ainsi, sans même évoquer les autres impacts de la pêche, comme l'atteinte de certains écosystèmes profonds [39] par les chaluts de fond (les "*racleurs d'océan*" [40]) ou sa contribution à la pollution au plastique du milieu marin par l'usure ou la perte du matériel de pêche [41]. Il apparaît que l'érosion de la biodiversité et de la biomasse des ressources halieutiques par la seule pression de pêche soient des raisons suffisantes à l'élaboration de systèmes de régulation des pêches avec pour toile de fond, la mise en garde de Rousseau, "*Vous êtes perdus si vous oubliez que les fruits sont à tous, et que la terre n'est à personne !*" [42].

1.2.2 La gestion des pêches dans l'Union Européenne

Dès les années 1950, en parallèle à ces constatations de l'épuisement des ressources marines, a émergé une prise de conscience collective de leur finitude [43]. De cette prise de conscience est née la nécessité d'introduire sur la scène internationale, notamment au travers des conférences des Nations Unies, la question de la conservation des ressources halieutiques appelant la mise en place de mesures de gestion afin d'éviter une crise de la pêche mondiale [44, 45].

Ces conférences successives (Rome en 1955 [46], Genève en 1958 [47], Stockholm en 1972 [48], etc.) ont permis de rassembler la plupart des Etats du monde autour d'objectifs communs pour une exploitation plus vertueuse à la fois économiquement et environnementalement. C'est par exemple en 1987 lors de la publication du rapport de la commission de Brundtland, qu'a été émise la nécessité de la mise en place de mesures de gestion visant à assurer la durabilité de l'exploitation des ressources naturelles. Dans ce rapport, la notion de durabilité est définie comme la capacité à "*satisfaire des besoins des générations présentes sans compromettre la capacité des générations futures à satisfaire leur propres besoins*" [49].

Concernant les ressources marines, la trajectoire menant à l'objectif de la durabilité a été jalonnée de sous-objectifs comme par exemple, la limitation de l'exploitation des stocks halieutiques à un rendement maximal durable [50], la réduction du rythme d'appauvrissement de la diversité biologique [51], la restauration de certains stocks [51] ou encore la création de zones marines protégées incarnée par l'objectif "30 × 30" de la COP 15 sur la biodiversité visant à "*protéger au moins 30% des terres et 30% des mers de la planète à l'horizon 2030*" [52].

A l'issue de ces conférences, les Etats s'engageant sur ces objectifs conservent leur

liberté dans le choix des actions et des stratégies de gestion à adopter dans le but de les atteindre. Ainsi, concernant l'Union Européenne (UE), face à une ressource sans frontière, le choix a été fait de rassembler les pays membres au travers d'une Politique Commune des Pêches (PCP) [53] effective depuis 1983. La PCP, issue de concertations entre le Conseil de l'Union Européenne et le Parlement Européen, fixe à l'échelle de l'UE "*les conditions de conservation des ressources halieutiques et de gestion des pêcheries et des flottes qui exploitent ces ressources*" [54]. La PCP tend à harmoniser les stratégies de gestion et à mutualiser les efforts de chacun des Etats membres.

1.2.3 La recherche scientifique au service de la gestion

L'un des enjeux de la PCP réside dans sa capacité à concilier des objectifs parfois antagonistes comme des objectifs économiques, sociaux et environnementaux. A cette fin, la recherche scientifique joue le rôle de points d'appuis et d'aide à la prise de décision. En effet, derrière la gestion des pêches se trouve une diversité d'axes de recherche concernant par exemple, l'évaluation de l'état des stocks ou des écosystèmes exploités et proposant des recommandations en terme de quotas de pêche, la compréhension des mécanismes intervenant dans la dynamique des ressources naturelles et des pêcheries pour une meilleure estimation de l'effort et de la pression de pêche, l'analyse de projections de modèle de pêche pour anticiper les réponses de la mise en place de différents scénarios de gestion ou d'autres forçages tel que le changement climatique.

Pour les acteurs de la recherche, l'enjeu derrière cette pluridisciplinarité consiste à rendre accessible des résultats de travaux de recherche, bien souvent aussi complexes que les mécanismes qu'ils tentent d'analyser. Cette accessibilité des résultats scientifiques peut être rendue possible par exemple par l'utilisation d'indicateurs pouvant être définis comme "*des outils qui permettent de donner une description simple, quantitative ou qualitative, de phénomènes ou de systèmes complexes*" [55] afin de "*favoriser la communication, la transparence, l'efficacité et la responsabilisation pour ce qui concerne l'aménagement des ressources naturelles*" [56].

L'un des besoins nécessaire à la conduite de la PCP est la constitution d'un jeu de données relatif à l'activité halieutique à l'échelle de l'Union Européenne. Ce besoin s'est traduit par l'intégration de la question de la collecte des données dans le "*cadre communautaire pour la collecte et la gestion des données nécessaires à la conduite de la PCP*" instauré en 2000 au travers de la DCR ("*Data Collection Regulation*"), devenue en 2008 la

DCF ("*Data Collection Framework*"), puis en 2016 la DC-MAP ("*Data Collection Multi Annual Program*"). Sans entrer dans le détail, ces règlements successifs imposent aux pays membres de l'UE l'élaboration de programmes de collectes annuels de données sur les navires de pêche, leurs activités, leurs captures et la santé économique des filières de la pêche.

En France, outre l'acquisition de données par le déploiement de campagnes scientifiques, ces règlements concernent également le recueil et la valorisation d'un certain nombre de données dites commerciales liées à la pêche parmi lesquelles se retrouvent :

- Des données de ventes des débarquements, correspondant au volume et à la valeur des ventes réalisées pour chaque navire de pêche.
- Des données déclaratives (*Log Book*), correspondant aux déclarations des patrons pêcheurs concernant leurs captures et les engins utilisés par jours de mer, par espèce et par rectangle statistique (localisation).
- Des données de géolocalisation, correspondant au suivi et à la surveillance des navires de pêche par des balises satellites VMS ("*Vessel Monitoring System*").

1.3 Du suivi des pêcheurs à l'analyse de leur comportement

Le recueil de ces données commerciales a fourni aux acteurs de la recherche une matière première permettant d'interroger une facette des pêcheurs longtemps restée peu considérée, leur comportement en mer. La compréhension de ce comportement revient à interroger la nature des activités d'un pêcheur et leurs successions dans le temps entre son départ et son retour au port. Une analogie du comportement d'un pêcheur peut consister à le considérer comme celui d'un prédateur "animal" à la recherche de proies. Sous cette abstraction, le comportement individuel d'un navire de pêche peut être appréhendé et décrit selon des concepts éthologiques, décomposant le comportement de prédation en une succession de trois phases [57], une phase de poursuite correspondant à la recherche, à l'identification et à la prise en chasse d'une proie d'intérêt, une phase de capture correspondant au déploiement de l'engin de pêche autour de la proie et à son prélèvement, et enfin, une phase de consommation correspondant dans ce cas, au traitement des prises à bord du navire via leur tri, leur préparation et leur conditionnement.

Une observation directe d'un navire de pêche, c'est-à-dire à l'oeil nu par l'intermé-

diaire d'un observateur à bord, ou indirect, c'est-à-dire au travers d'un dispositif de capture photographique ou vidéo permet d'identifier la succession de ces différentes phases de poursuite, de capture ou de consommation, ne serait-ce que par le déploiement ou le relèvement de l'engin de pêche marquant le début et la fin de la phase de capture. Les données commerciales ne fournissent pas ce type de données d'observation pouvant cependant être obtenues ponctuellement au travers de programme faisant intervenir des observateurs scientifiques comme le permet le programme Obsmer [58]. Toutefois, les données commerciales fournissent un aspect clef du comportement des navires de pêche, leurs trajectoires rendues accessibles au travers des données de suivi VMS.

1.3.1 La donnée VMS

Le système de surveillance VMS ("*Vessel Monitoring System*") est au sein de l'UE le principal système de surveillance par satellite fournissant aux autorités de pêche des données sur la localisation des navires de pêche. Il s'agit d'une donnée de géolocalisation par satellite (GPS) produite par un système embarqué à bord de chaque navire de pêche commerciale. Mis en place en 1992 pour les navires de plus de 24m, l'embarquement de balises VMS a progressivement été rendu obligatoire par l'UE aux navires de plus petite taille, pour les plus de 15m en 2005, et pour les plus de 12m en 2012 .

Initialement mis en place afin de garantir le respect des règles relatives aux lieux et aux périodes de pêche, son enregistrement à raison d'un point toutes les heures en moyenne a offert la possibilité de reconstruire les trajectoires des navires de pêche ouvrant la porte de leurs analyses.

1.3.2 De la trajectoire au comportement individuel

Qu'il s'agisse d'un animal ou d'un navire de pêche, lorsque l'unique matériel disponible est l'échantillonnage d'une trajectoire, il ne s'agit plus d'interroger le comportement de l'individu, mais bel et bien celui de sa trajectoire. A la faveur de la démocratisation et de la miniaturisation des traceurs GPS, offrant la possibilité de suivre un grand nombre d'individus sans leur imposer la présence d'un observateur, de nombreux travaux, notamment d'écologie, ont pu s'intéresser à l'analyse d'une vaste collection de trajectoires d'animaux en tout genre [59, 60], et au développement de statistiques relatives à leur déplacement dans le temps et l'espace [61].

Concernant les trajectoires des navires de pêche, définies comme des séries temporelles discrètes $X_{t \in \mathbb{N}}$ où chaque pas de temps t est associé à deux variables x_t et y_t , respectivement la longitude et la latitude, la vitesse moyenne $v(t) = \frac{X_{t+1} - X_t}{(t+1) - t}$, définie par la distance parcourue entre deux pas de temps successifs, apparaît comme un paramètre clef de l'identification des différentes phases comportementales [62, 63].

Par exemple, lorsque les navires de pêche sont des chalutiers, les histogrammes des vitesses (Fig 1.5) font ressortir plus ou moins clairement selon le type de chalut, deux régimes de vitesse, un lent (≤ 4.5 *noeuds*) et un rapide (> 4.5 *noeuds*) interprétés respectivement comme des vitesses relatives à des périodes de pêche effective (engin déployé) et des périodes de route ou de recherche.

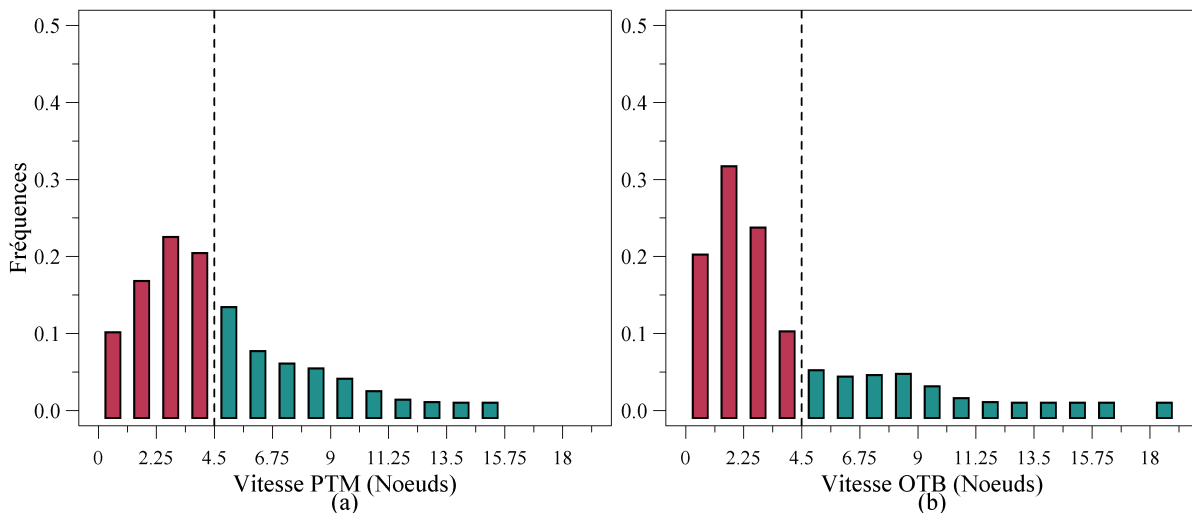


FIGURE 1.5 – Exemple d’histogrammes des vitesses ($v(t)$) de deux flottilles de chalutiers (a) en boeufs et (b) de fond, présentant deux régimes de vitesse (seuil à 4.5 noeuds (*nd*)).

Les trajectoires individuelles peuvent ainsi être décomposées en une suite de segments de pêche effective, de recherche ou de déplacement (Fig 1.6)² faisant de la vitesse moyenne un proxy de l’activité de pêche à mettre en relation avec les données déclaratives des pêcheurs. La combinaison de ces deux sources d’information se traduit par un gain de précision et de compréhension de la spatialisation des captures et par voie de conséquence, une meilleure estimation spatiale et temporelle de l’effort de pêche, indispensable au suivi et à l’amélioration des stratégies de gestion [64].

2. Pour des raisons de confidentialité, les trajectoires des navires de pêche issues des données VMS seront présentées dans ce manuscrit au travers de figures sans axes gradués ni projection des traits de côtes.

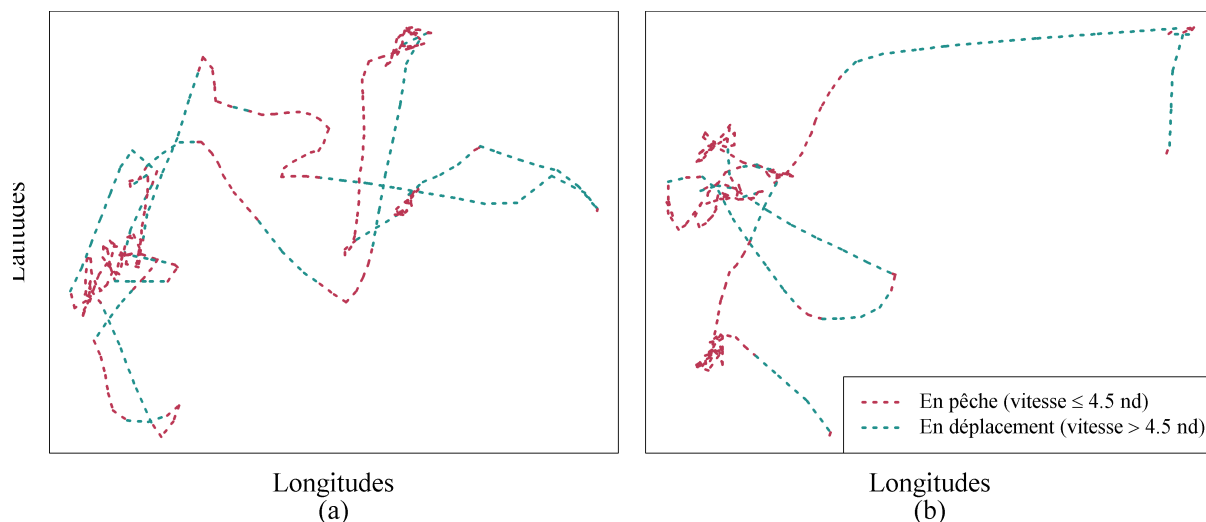


FIGURE 1.6 – Illustration de la segmentation d’une trajectoire en périodes de pêche et de déplacement selon la vitesse pour deux types de chalutiers (a) en boeufs, (b) de fond.

1.3.3 Du comportement individuel au comportement deux à deux

Les approches reposant sur l’analyse individuelle des trajectoires ignorent la présence éventuelle d’autres pêcheurs et ainsi les potentielles relations les unissant, et par la même occasion, les comportements synergiques ou antagonistes qui, s’ils existent, peuvent influencer positivement ou négativement la capture de la pêche. Il apparaît alors pertinent que les pêcheurs ne soient plus considérés individuellement en tant qu’individus isolés mais en tant qu’un ensemble d’individus regroupés en flottille. Dans ce cas, l’analogie entre pêcheurs et prédateurs apparaît d’autant plus pertinente lorsque la définition de la flottille repose sur un critère tel que l’engin de pêche, supposant la convoitise de la même ressource/proie. Considérer une population de navires convoitant les mêmes proies, rend probable l’existence de relations de collaboration ou de compétition entre les différents membres de la flottille, en d’autres termes, l’existence d’une forme de comportements sociaux.

Le prolongement de l’analyse du comportement individuel vers l’analyse du comportement collectif a été marqué par un travail proposé par Joo et al [65] en 2021 comme un point d’étape intermédiaire en proposant une approche interrogeant le comportement deux à deux des pêcheurs. Dans ce travail, les auteurs s’appuient sur la conception d’un objet élémentaire, la dyade, définie par un couple de trajectoires simultanément en mer

(Fig 1.7). Au travers des dyades, les auteurs proposent d’interroger le comportement d’un navire au regard du comportement d’un second en évaluant la cohésion de leur trajectoires respectives.

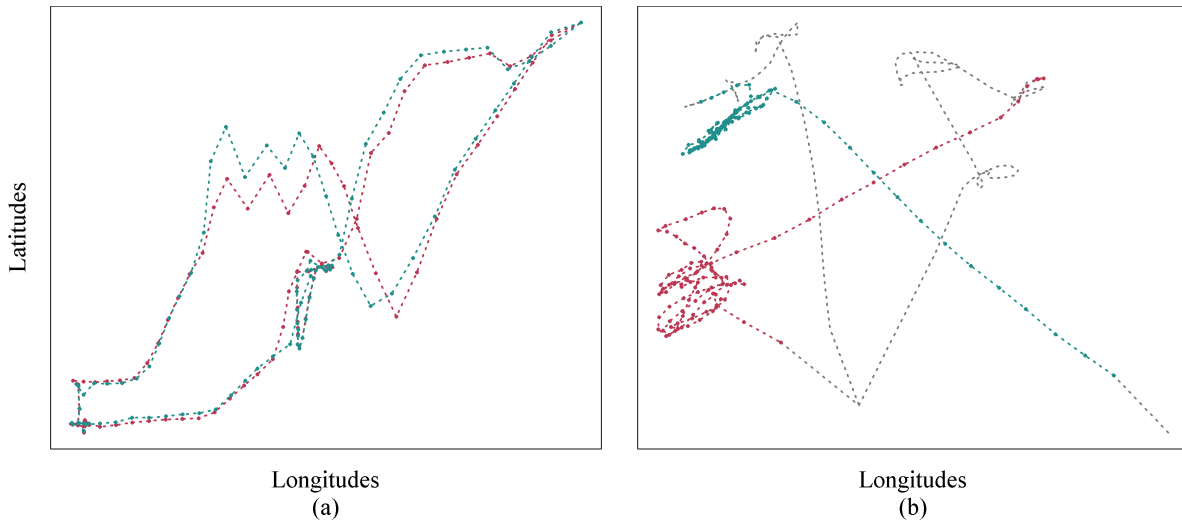


FIGURE 1.7 – Illustration de deux couples de trajectoires formant une dyade (a) à forte cohésion (collaboration) (b) à faible cohésion (croisement).

Cette évaluation de la cohésion d’une dyade dans le temps et dans l’espace est effectuée à l’aide d’un ensemble de métriques. Une approche par classification permet alors de discriminer les dyades selon l’intensité de leur cohésion en différents groupes chacun associé à un type de comportement deux à deux allant de la collaboration durable (Fig 1.7a) au simple croisement mer (Fig 1.7b). Nous reviendrons plus en détail sur la conceptualisation des dyades et l’identification des comportements deux à deux au cours du chapitre suivant.

1.3.4 Du comportement deux à deux au comportement N à N

L’identification des relations dyadiques offre la possibilité de décomposer un ensemble de trajectoires en un ensemble de dyades. Faisant suite à cette décomposition, Joo et al ont proposé une approche permettant de mettre en perspective l’ensemble des relations dyadiques les unes par rapport aux autres sous la forme d’un graphe (Fig 1.8). Le travail décrit dans ce manuscrit propose d’interroger le comportement collectif d’un groupe de navire par l’analyse des caractéristiques topologiques du graphe de leurs relations dyadiques.

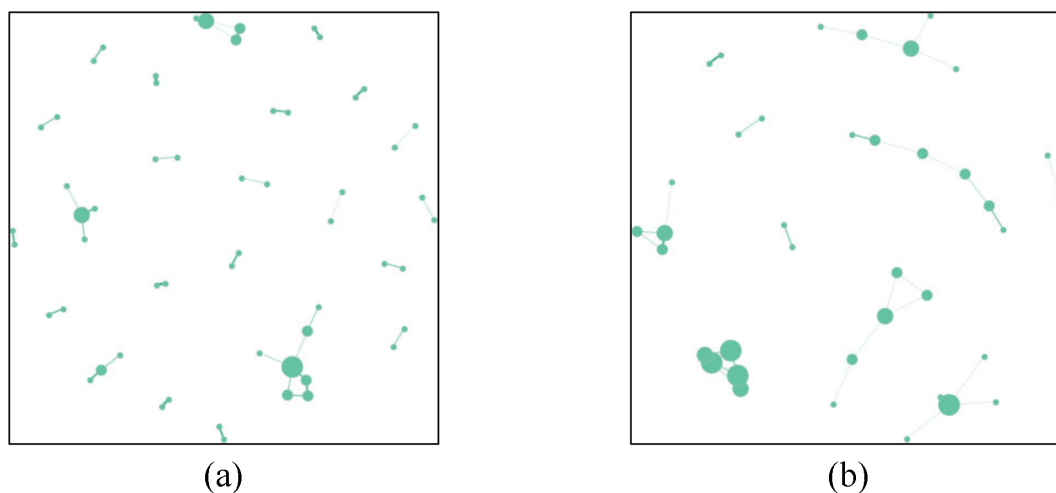


FIGURE 1.8 – Exemple de graphes des relations dyadiques de collaboration entre un groupe de navires. Chaque noeud représente un navire dont la taille est proportionnelle au nombre de dyades dans lesquelles il est impliqué. (D’après Joo et al [65]).

Une hypothèse souvent exprimée par défaut en halieutique considère les navires de pêche comme des entités indépendantes les unes des autres, sous-tendant une hypothèse d’indépendances statistiques. En d’autres termes, si les captures des différents navires de pêche sont considérées comme la réalisation de variables aléatoires, ces variables sont par hypothèse considérées comme indépendantes. Or, la validation de cette hypothèse d’indépendance peut être requise lors de l’utilisation des données de captures dans certaines analyses statistiques intervenant par exemple dans les méthodes d’évaluations des stocks ou d’estimation de l’effort de pêche. De même certains modèles de simulation de trajectoires nécessitent, par exemple à des fins de calibration, l’utilisation de trajectoires réelles indépendantes.

L’existence de comportements de collaboration deux à deux soulève l’existence de trajectoires deux à deux non indépendantes et devrait induire une non indépendance mutuelle des trajectoires à l’échelle d’une flottille. Toutefois, la non existence de ces comportements de collaboration est insuffisante pour conclure à la non indépendance d’un ensemble de navire à l’échelle d’une flottille. Dans ce manuscrit, l’analyse des graphes des relations dyadiques sera centrée sur la question de leur indépendance apparaissant ainsi comme la première marche menant à la caractérisation du comportement collectif. Paradoxalement, la validation de cette hypothèse d’indépendance, d’une certaine manière synonyme d’une absence de comportement collectif, ferait de cette marche à la fois la première et la dernière.

1.4 Le graphe, un objet mathématique

La conceptualisation des graphes et leur analyse sont contenues dans une discipline connue sous le nom de théorie des graphes [66, 67]. Dans cette section, il est question de fournir les notions fondamentales issues de la théorie des graphes afin de rendre accessibles les développements qui suivront. Dans un premier temps, au travers d'exemples historiques, nous verrons qu'il n'existe pas un graphe mais des graphes, correspondant à différentes façons de représenter des relations unissant un ensemble d'entités. Dans un second temps, nous verrons qu'au-delà d'une représentation de phénomènes réels, des graphes peuvent être entièrement construits par simulation par l'intermédiaire de modèles de graphes dont nous présenterons les représentants les plus emblématiques.

1.4.1 Définition et formalisation d'un graphe

Le terme de "graphe" est inféodé aux mathématiques mais peut être substitué dans la littérature par le terme de "réseau" notamment en écologie ou en sciences sociales. Formellement, un graphe G est défini par un ensemble $V(G) = \{v_1, v_2, \dots, v_n\}$ d'éléments appelé noeuds (ou *sommets*, "**V**ertices" en anglais) et par un ensemble $E(G) = \{e_1, e_2, \dots, e_m\}$ d'éléments appelé arêtes ("**E**dges" en anglais), selon $G = (V, E)$. En l'absence d'arêtes, $E(G) = \emptyset$, le graphe est dit vide et n'est constitué que de noeuds dits isolés. Dans le cas contraire, si $E(G) \neq \emptyset$, alors pour chaque élément $e = \{v_i, v_j\} \in E(G)$ il est dit que e est l'arête de G d'extrémités v_i et v_j , ou bien que e joint v_i et v_j . Les noeuds v_i et v_j sont alors dits adjacents dans G .

$$A = \begin{matrix} & v_1 & \cdots & v_j & \cdots & v_N \\ \begin{matrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_N \end{matrix} & \begin{bmatrix} a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,N} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,N} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{N,1} & \cdots & a_{N,j} & \cdots & a_{N,N} \end{bmatrix} \end{matrix} \quad \text{Avec} \quad a(i, j) = \begin{cases} 1 & \text{si } (v_i, v_j) \in E \\ 0 & \text{sinon} \end{cases} \quad (1.1)$$

Mathématiquement, un graphe peut être formalisé par une matrice d'adjacence notée A , contenant autant de lignes et de colonnes que le graphe contient de noeuds. Ainsi, chaque ligne i ou colonne j correspond à un noeud v_i ou v_j (Eq 1.1). Chaque élément de la matrice a_{ij} permet de décrire les relations d'adjacences entre les noeuds v_i et v_j . Sous

sa forme la plus "simple", la matrice d'adjacence est binaire selon $a_{ij} = 1$ si v_i et v_j sont adjacents ou $a_{ij} = 0$ sinon.

Certains phénomènes se traduisent naturellement sous la forme d'un graphe, comme par exemple, un arbre généalogique où les arêtes traduisent les liens de filiation entre les individus [68], l'organigramme d'une entreprise traduisant les relations hiérarchiques entre les employés [69] ou encore le réseau trophique d'un écosystème traduisant les relations de prédatons entre les espèces ou les groupes trophiques [70]. Ces différents exemples de graphes paraissent familiers, du fait de la nature ordinaire et commune de la relation représentée.

D'autres graphes peuvent apparaître plus abstraits, compte tenu de la relation plus "subtile" ou moins tangible qu'ils représentent. C'est le cas par exemple de la communauté mondiale des chercheurs pouvant être représentée sous la forme d'un graphe où les arêtes traduisent les relations de citations [71], d'un ensemble de protéines traduisant l'existence ou le partage d'un même site catalytique [72], de la propagation d'une épidémie en population générale [73], ou encore le graphe des espèces pollinisatrices et des espèces des plantes pollinisées [74].

Afin de représenter au mieux le phénomène décrit par un graphe, plusieurs types de graphes sont apparus afin notamment de capturer au mieux, les modalités de la relation entre deux entités (Fig 1.9).

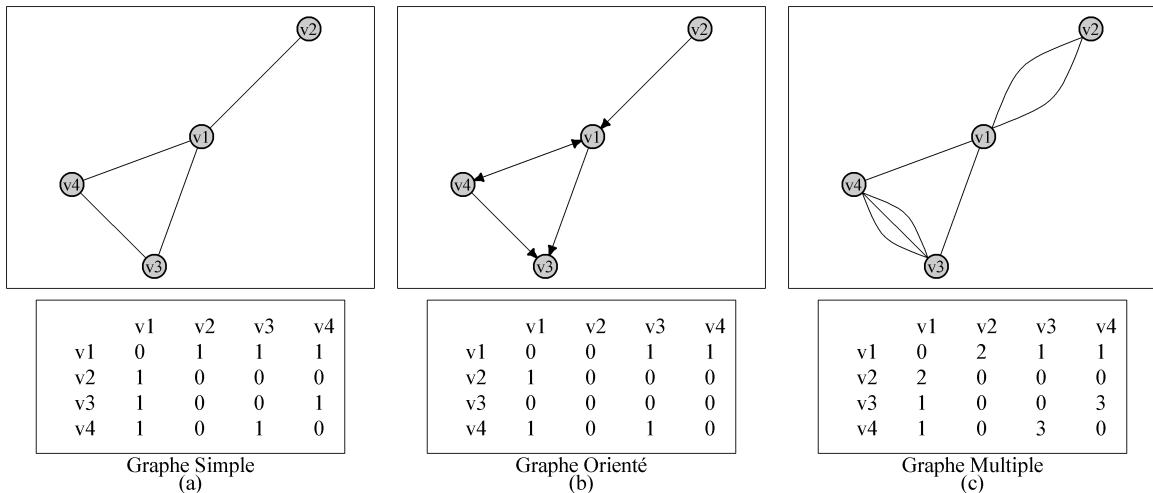


FIGURE 1.9 – Les différents types de graphes (a) simple, (b) orienté et (c) multiple.

Le cas des graphes simples :

Les graphes simples (Fig 1.9a) sont caractérisés par le fait qu'un couple de noeuds ne peut être relié que par une unique arête. Ainsi, les graphes simples permettent seulement de mettre en évidence l'existence d'une relation entre 2 noeuds. Dans le cas d'un graphe simple, la matrice d'adjacence est binaire et symétrique. Dans la littérature, l'un des plus célèbres graphes simples est probablement celui du "Karate Club" de Zachary [75] paru en 1977 (Fig 1.10). La notoriété de ce graphe ne vient pas du phénomène qu'il représente, il s'agit d'un simple réseau social, mais de l'utilisation de ce graphe comme le support de la compréhension d'un phénomène de rupture au sein d'un groupe.

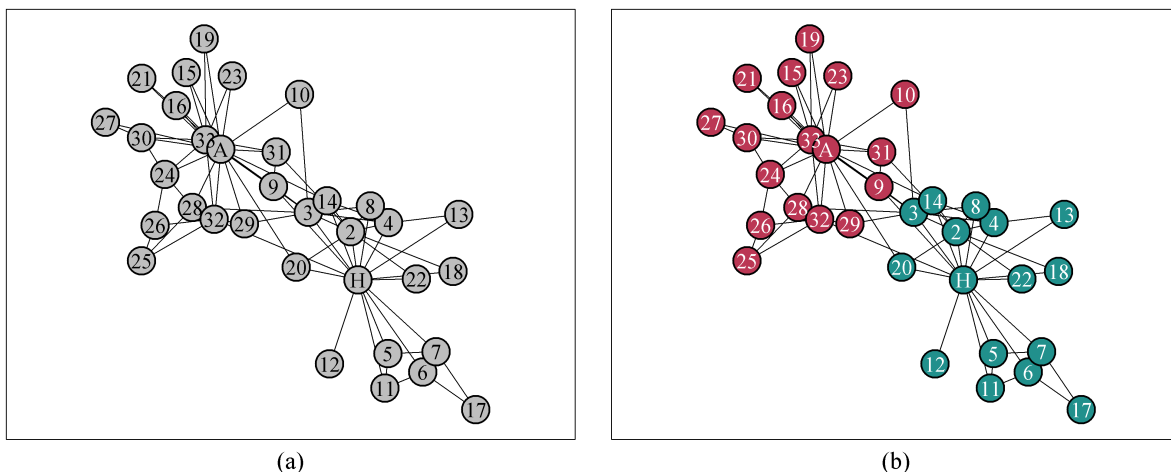


FIGURE 1.10 – Représentation du graphe (a) des relations sociales du club de karate de Zachary [75] et (b) de ces deux communautés.

Dans un club de karaté une dispute éclate entre le directeur du club "M.A" et un professeur à mi-temps "M.H" à propos des tarifs d'inscription au club que le second juge trop faibles. Le conflit initialement larvé, s'envenimât pendant près de 2 ans aboutissant à la séparation du club en deux clubs distincts, celui de M.A et celui de M.H, chacun emmenant avec lui une partie des membres du club.

Peu de temps avant la fission définitive du club, Zachary a construit le réseau social de ses adhérents sur la base des relations qu'ils entretenaient en dehors des activités du club. Pour chaque membre (les noeuds), une arête avec un autre membre implique l'existence d'une relation amicale caractérisée par des rencontres régulières en dehors des activités liées à celle du club (Fig 1.10a). Sous cette forme, il s'agit d'un simple réseau social traduisant des liens d'amitié.

Zachary considéra ce graphe, comme le support d'un réseau d'information, et interrogea la manière dont un "flux d'information" diffuse au travers de ce réseau, et à la manière dont chaque membre perçoit une information émise. Sur la base de ces flux d'informations, le réseau social a été découpé en deux groupes ou communautés, définissant des ensembles de noeuds dont l'information se propage "mieux" entre membres d'un même groupe qu'entre membres de groupes différents (Fig 1.10b). Les membres des deux groupes ainsi obtenus sont différenciés par des couleurs de noeuds différentes (Fig 1.10b).

La notoriété du graphe de Zachary vient du fait, que lorsque la fission eut lieu, la séparation effective des membres en deux groupes correspondait à la séparation des membres obtenue par l'analyse du flux d'information au travers du graphe. En d'autres termes, que la structure du graphe, son organisation ou en termes de graphe sa topologie, c'est-à-dire la manière dont les arêtes ou les noeuds sont organisés les uns par rapport aux autres, rendait compte du conflit interne au club de karaté, et projetait (pour ne pas dire prédire) la séparation des membres du club en deux groupes. Seul un membre, le noeud n°9, situé à l'interface entre les deux groupes s'est vu être mal classé.

Le cas des graphes multiples :

Les graphes multiples (Fig 1.9c) par opposition aux graphes simples présentent la particularité pour un couple de noeuds, de pouvoir être reliés, théoriquement, par un nombre infini d'arêtes. Le recours aux graphes multiples offre deux possibilités en fonction de ce que représente la multiplicité des arêtes.

La première possibilité est d'utiliser une arête multiple comme la traduction d'une répétition d'une même relation entre deux noeuds. Dans ce cas, la matrice d'adjacence toujours symétrique, n'est cette fois-ci plus binaire où chacun de ses éléments $a_{i,j}$ correspond au nombre d'arêtes entre le noeuds v_i et v_j . La seconde possibilité est d'utiliser une arête multipliée, afin de distinguer l'existence de relations de natures différentes entre un même couple de noeuds.

A l'origine des graphes multiples se trouve un graphe bien plus ancien que celui du club de karaté de Zachary, le graphe d'Euler (Fig 1.11b) ayant permis de prouver en 1736, l'impossibilité de résoudre le "problème des 7 ponts de Königsberg" [76], qui est énoncé comme suit : *"Dans le ville de Königsberg se trouve une île entre deux bras de rivière. Il y a au total 7 ponts permettant de franchir les différents bras de la rivière. La question est alors, existe t-il un chemin permettant de traverser tous les ponts seulement une fois ?"*

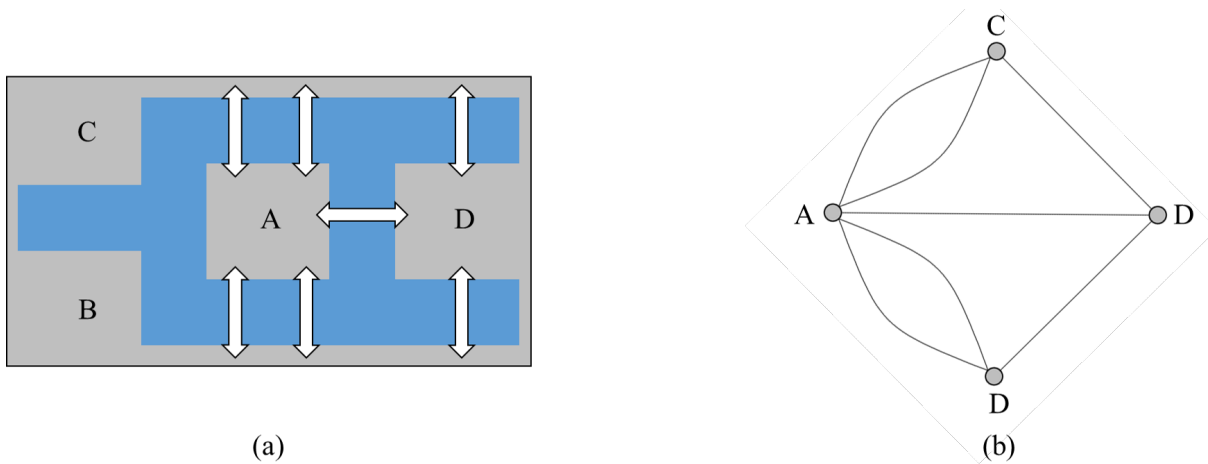


FIGURE 1.11 – Représentation schématique de la ville de Königsberg et de ses sept ponts (a) et (b) sa traduction en graphe d'Euler.

Pour répondre à ce problème, Euler l'a traduit sous la forme d'un graphe. Dans un premier temps, il a défini toutes les zones séparées par l'écoulement de la rivière comme autant de points pouvant être reliés par l'intermédiaire d'un ou de plusieurs ponts (Fig 1.11a), l'île décrite dans l'énoncé est alors le point "A", les rives gauches et droites sont les points "B" et "C" et enfin le point "D" pour le reste. Le problème est ainsi réduit sous la forme d'une liste recensant toutes les possibilités pour rejoindre les différents points $\{A \leftrightarrow B; A \leftrightarrow B; A \leftrightarrow C; A \leftrightarrow C; A \leftrightarrow D; D \leftrightarrow C; D \leftrightarrow B\}$ qui n'est ni plus ni moins qu'une liste d'arêtes (qui est une possibilité alternative à la matrice d'adjacence de formaliser un graphe).

Dans un second temps, il a construit le graphe mettant en relation ces différents points (les noeuds) au regard des ponts (les arêtes) permettant de passer de l'un à l'autre. Du fait qu'il existe 2 ponts permettant de passer du point "A" au point "B", ou du point "A" au point "C", deux arêtes (une arête multiple) sont représentées entre ces deux couples de noeuds. Dans ce cas, le graphes d'Euler, utilise les arêtes multiples pour spécifier l'existante de 2 relations différentes (deux ponts différents) entre un même couple de noeuds.

Non content de montrer qu'il était impossible de traverser seulement une fois tous les ponts de Königsberg, Euler présenta une démonstration générale permettant de déterminer s'il existe un chemin traversant tous les ponts seulement une fois, quelle que soit la configuration de départ, en terme de nombre de points $\{A, B, C, \dots\}$ et de nombre de ponts reliant ces points, selon :

1. *S'il existe plus de deux régions, présentant un nombre impair de ponts, aucun chemin ne permet de traverser tous les ponts seulement une fois.*(C'est le cas des 7 ponts de Königsberg).
2. *Si, cependant, seulement 2 régions présentent un nombre impair de ponts, le chemin existe si le point de départ est une de ces deux régions.*
3. *Enfin, si aucune région ne présente un nombre impair de ponts, un chemin existe quelle que soit la région de départ.*

Au travers de ce problème, et de ce graphe multiple, en reformulant le problème sous une forme abstraite, une liste de points reliés entre eux, Euler posa les bases de ce qui deviendra plus tard, la théorie des graphes, dont sa démonstration en constitue le premier théorème [77].

Le cas des graphes orientés :

Jusqu'ici, les graphes qui ont été présentés sont des graphes dits "non-orientés", les différenciant des graphes dits "orientés" (Fig 1.9b) dont la particularité est de pouvoir spécifier, s'il existe, le sens (ou la direction) de la relation entre 2 noeuds. Les graphes orientés offrent ainsi la possibilité de représenter des relations unilatérales entre un noeud émetteur, qui est à l'origine de la relation, et un noeud cible qui la reçoit. En accord avec la notion de noeud émetteur et de noeud cible, les arêtes orientées sont appelées "arcs" afin de les distinguer des arêtes, elles, considérées par défaut non orientées. Dans ce cas, la matrice d'adjacence est cette fois asymétrique, permettant de distinguer les noeuds émetteurs des noeuds cibles. Contrairement aux graphes simples et aux graphes multiples, les dyades n'étant pas des relations unilatérales, aucun des graphes de dyades construit dans le chapitre précédent, ne permet d'illustrer des arcs entre les navires ou les marées. L'illustration d'un graphe orienté se fera alors uniquement au travers d'un graphe issu de la littérature, le graphe de Sampson [78].

En 1968, S.F Sampson était doctorant en sociologie à l'Université de Cornell où il s'intéressait aux liens sociaux existant au sein d'une petite collectivité de moines résidents au sein d'un monastère de la Nouvelle-Angleterre. Il intégra alors le monastère en tant qu'observateur pendant près d'un an, période durant laquelle il eut l'occasion de s'entretenir à plusieurs reprises avec les différents moines.

Pendant son séjour, il s'intéressa particulièrement à un petit groupe de 18 jeunes moines qualifiés de "novices", ayant récemment intégré le monastère afin de parfaire leur

éducation religieuse avant de choisir un Ordre. Au cours de son séjour, il réalisa quatre séries d'entretiens avec chacun des novices afin d'interroger les relations que chacun d'entre eux entretenait avec les autres. Pour ce faire Sampson définit 4 types de relations présentant chacune un pendant positif et un autre négatif selon : "Amicale ou non amicale", "Estime positive ou négative", "Influence positive ou négative", "Louanges ou Blâme". Chaque moine devait alors citer pour chacune de ces relations, son "top 3" parmi les autres moines. Par exemple le moine "A", cite les moines "B" en 1, "C" en 2 et "D" en 3 pour les relations amicales et les moines "E" en 1, "F" en 2 et "G" en 3 pour les relations non amicales et ainsi de suite pour toutes les relations. Il s'agit donc bien ici de relations unilatérales d'écrites à l'aide d'arcs en guise d'arêtes.

Peu de temps après la dernière série d'entretiens et le départ de Sampson du monastère, une violente crise ("*crisis in a cloister*") éclatât aboutissant à une série d'expulsions et de départs volontaires d'une partie des novices. Le déroulement de cette crise peut être illustré par une série de graphes construit a posteriori sur la base des observations de Sampson [79] (Fig 1.12). Un conflit éclatât entre le groupe défini par Sampson des "Jeunes Turcs", opposés à certaines règles régissant la vie du monastère, et le groupe défini par Sampson de l'"Opposition Loyale" défendant les règles en vigueur.

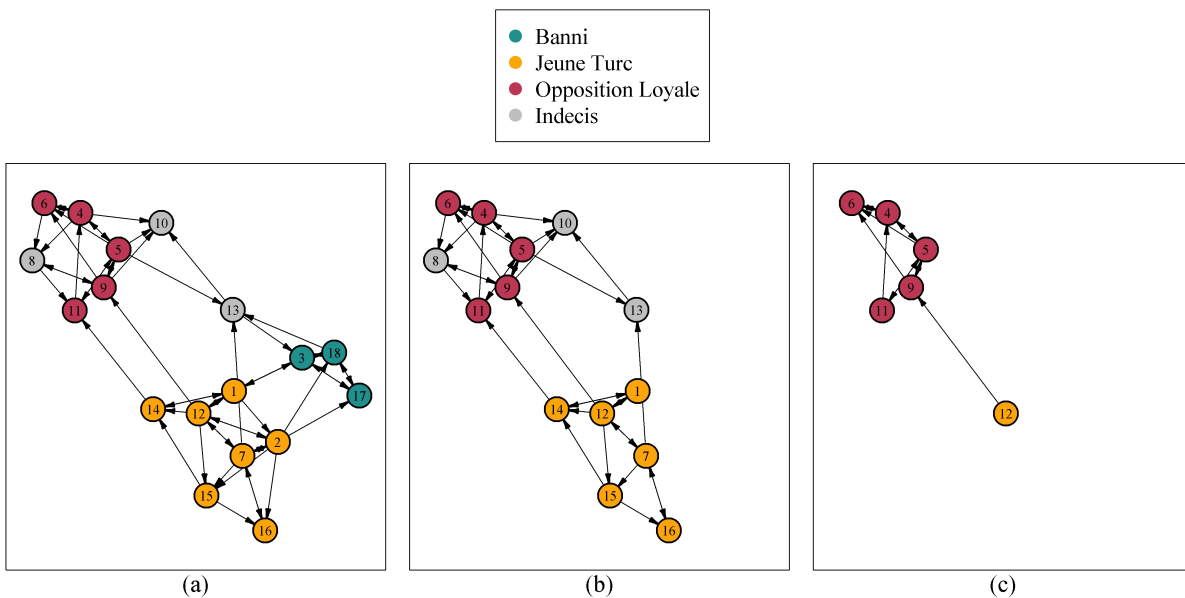


FIGURE 1.12 – Illustration des graphes de Sampson des relations sociales entre les moines novices d'un monastère (a) avant la crise, (b) après l'exclusion d'une partie des novices et (c) après la vague de départs volontaires.

La décision fut prise d'exclure quatre moines (2, 3, 17 et 18) (Fig 1.12a), les 3 derniers pour leur "immaturité" et le premier pour son caractère "trop indépendant, interrogateur et arrogant" et considéré par Sampson comme le leader du groupe des Jeunes Turcs. A la suite de cette exclusion, une succession de départs volontaires a eu lieu, tout d'abord le départ du novice 1 (le "second leader" du groupe des Jeunes Turcs), suivi quelques jours plus tard des novices 7, 14, 15 et 16 suivi à leur tour des moines 8, 10 et 13 (Fig 1.12b) ne laissant derrière eux que le groupe des moines "Loyaux" et un unique "Jeune Turc" (Fig 1.12c).

Au delà du caractère tragique que revêt l'histoire de ces graphes, la forte documentation qui les entoure, issue des observations de Sampson, a fait de ces graphes un cas de référence des graphes orientés retraçant l'émergence et la scission d'un groupe en une somme de communautés. Les graphes du monastère de Sampson sont devenus des incontournables de la théorie des graphes, notamment concernant le développement et la mise à l'épreuve d'analyses d'identification des communautés au sein d'un graphe, telle que l'approche des stochastic blockmodels [80]. Une question reste en suspens, pourquoi le moine 12 (Jeune Turc) est-il resté au sein de "l'Opposition Loyale" ?

Et les autres cas... :

Les graphes simples ou multiples, orientés ou non, représentent les principaux types de graphes rencontrés en théorie des graphes car permettant de traduire la majorité des phénomènes observés. Toutefois d'autres types de graphes existent, comme les graphes valués (ou pondérés) [81] ou les hypergraphes [82]. De plus il n'existe pas de frontières hermétiques entre ces différents types de graphes, offrant la possibilité de les combiner à loisir, où en théorie rien n'interdit l'existence simultanée d'arêtes et d'arcs, d'arêtes simples et multiples elles-mêmes orientées ou non.

Nous précisons dès à présent qu'il ne sera question dans ce manuscrit que de graphes simples non-orientés.

1.4.2 Des graphes empiriques aux modèles de graphes aléatoires :

Les quelques graphes présentés précédemment, qualifiés de graphes empiriques ou du monde réel ("*real-world graph*"), démontrent que les graphes sont plus que de simples outils de visualisation. Ils sont la résultante ou la traduction de phénomènes et de processus définissant l'existence d'une arête entre un couple de noeuds donné, et son absence entre

un autre. Comprendre, la structure d'un graphe, sa topologie, c'est-à-dire la manière dont les arêtes sont organisées les unes par rapport aux autres [83], c'est approcher les processus à l'origine du graphe.

C'est dans cette optique que sont apparus un certain nombre de modèles de graphes [84] permettant de générer des graphes dont l'existence d'une arête entre deux noeuds quelconque est définie par un processus ou une règle connue. Plusieurs modèles de graphe ont ainsi vu le jour, chacun défini par des processus génératifs différents visant à produire des graphes présentant des propriétés topologiques différentes. Un certain nombre de ces modèles ont été développés dans le but de mimer ou de reproduire des structures topologiques existantes (des graphes du monde réel). Les modèles de graphes sont dits aléatoires, lorsque le processus génératif des arêtes suit une loi de probabilité. Bien que le premier modèle de graphe aléatoire date des années 1950, il faut attendre la fin des années 1990 et le début des années 2000 pour que leur utilisation envahissent la littérature scientifique (Fig 1.13).

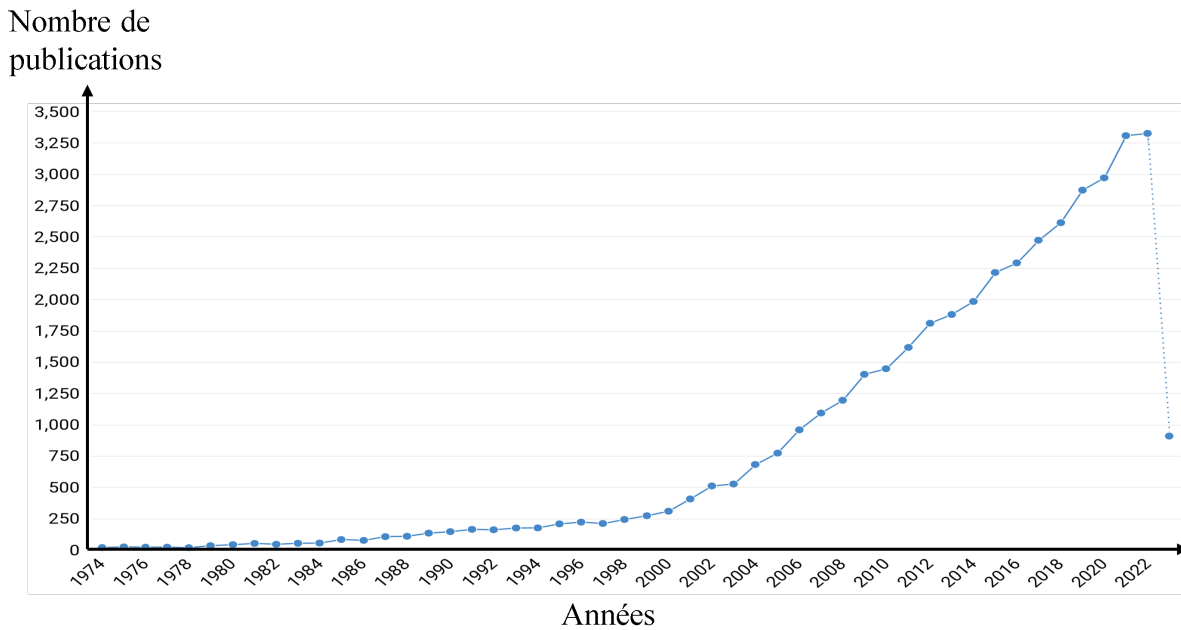


FIGURE 1.13 – Série chronologique du nombre de publications scientifiques contenant le mot clef "graph model" dans le titre ou le résumé de 1974 à 2022. (Données obtenues depuis la base de données Dimensions.ai fourni par Digital Science (<https://www.dimensions.ai>) consulté le 26 avril 2023)

Le modèle d'indépendance d'Erdős-Rényi (et Gilbert) :

Le modèle dit de Erdős-Rényi du nom de ses auteurs Paul Erdős et Alfred Rényi est considéré comme le premier modèle de graphes aléatoires paru en 1960 [85] et se définit comme suit : "*Dans le modèle de Erdős-Rényi $G_{ER}(N, m)$, un graphe est choisi aléatoirement selon une loi uniforme parmi la collection de tous les graphes à N noeuds et m arêtes*".

Nous précisons qu'un graphe à N noeuds et m arêtes est dit d'ordre N (nombre de noeuds) et de taille m (nombre d'arêtes). En considérant qu'un graphe d'ordre N peut contenir au maximum $N(N-1)/2$ arêtes, la définition du modèle de Erdős-Rényi peut être reformulée ainsi : "*Un graphe $G_{ER}(N, m)$ est une réalisation du modèle de Erdős-Rényi, si les m arêtes sont tirées aléatoirement et sans remise selon une loi uniforme parmi les $N(N-1)/2$ arêtes possibles*". Ainsi un graphe d'Erdős-Rényi $G_{ER}(20, 95)$ est un graphe d'ordre 20 noeuds et de taille 95 arêtes (Fig 1.14a) où chaque arête a une probabilité $p = \frac{m}{N(N-1)/2} = \frac{95}{190} = \frac{1}{2}$ d'être choisie parmi les $N(N-1)/2$ arêtes possibles.

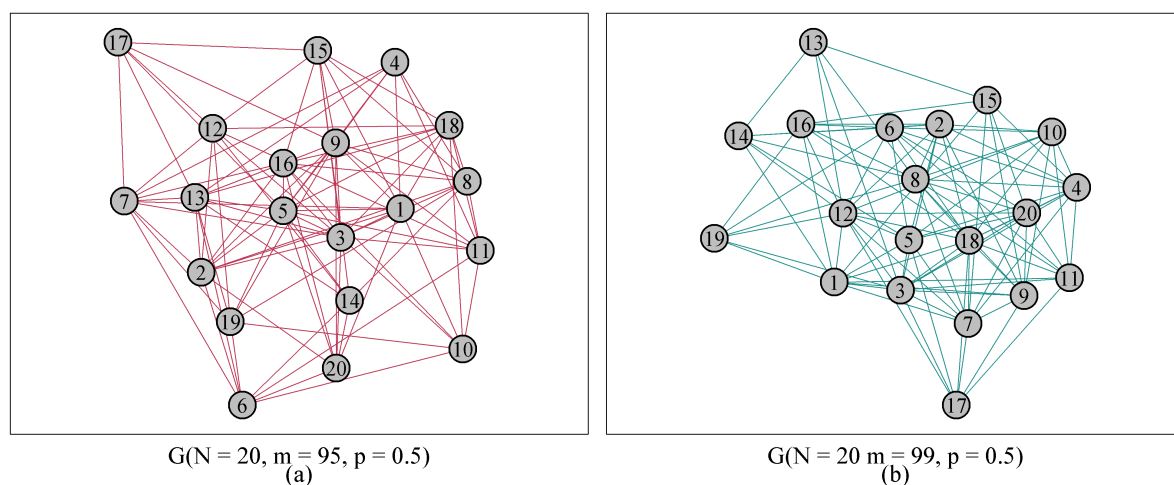


FIGURE 1.14 – Exemples de réalisations des deux modèles (a) d'Erdős-Rényi et (b) d'Erdős-Rényi-Gilbert

En réalité, un autre modèle très similaire à celui de Erdős et Rényi est paru en 1959 soit une année plus tôt par Edgar N. Gilbert [86] et se définit selon : "*Dans le modèle $G_{ER}(N, p)$, un graphe d'ordre N est construit en connectant aléatoirement chaque couple de noeuds avec une probabilité fixe p* " (Fig 1.14b). Selon ce modèle, la taille m du graphe n'est pas fixée et dépend de p , ainsi deux réalisations de ce modèle peuvent présenter une taille m différente (Fig 1.14).

Considérant la forte similarité entre ces deux modèles avec $p \approx \frac{2m}{N(N-1)}$ lorsque $N \rightarrow \infty$ [84] et les hasards de la publication scientifique favorisant le modèle publié par Erdős et Rényi, ces deux modèles sont souvent confondus et indistinctement retrouvés dans la littérature sous l'appellation de modèle d'Erdős-Rényi, bien que le modèle de Gilbert soit parfois mentionné en tant que modèle d'Erdős-Rényi-Gilbert [87]. Au delà de ces considérations nominatives, en tant que pionnier des modèles de graphes aléatoires, le modèle d'Erdős-Rényi a été richement étudié, notamment sous sa forme du modèle de Gilbert $G_{ER}(N, p)$ dont plusieurs propriétés mathématiques et topologiques ont pu être identifiées.

La première de ces propriétés est relative à sa dynamique topologique, c'est-à-dire, à la manière dont évolue la structure du graphe lorsque sa taille m augmente. En effet, bien que les graphes présentés précédemment soient considérés comme statiques, c'est-à-dire décrivant l'état d'un phénomène à un temps t ou intégré le long d'un temps t , un graphe peut aussi être vu un objet dynamique dans lequel les connections entre les noeuds peuvent être réversibles au cours du temps (des relations amicales entre les membres du club de karaté peuvent se faire ou se défaire, de même que de nouveaux ponts peuvent être construits ou que des anciens peuvent être détruits dans la ville de Königsberg).

Ainsi, dans le modèle $G_{ER}(N, p)$, en notant $\lambda = pN$, l'évolution de λ en fonction du paramètre p permet de définir des phases de transition dans l'évolution de la topologie des graphes (Fig 1.15) issues de la réalisation de ce modèle selon :

- Si $\lambda = 0$, soit ($p = 0$), le graphe ne contient aucune arête, le graphe est dit **vide**, constitué de noeuds isolés (Fig 1.15a)
- Si $\lambda < 1$, le graphe est constitué d'un ensemble de petites **composantes connexes** (groupe de noeuds liés entre eux) et ne possède presque sûrement aucune composante connexe de taille m supérieur à $O(\log(N))$ (Fig 1.15b).
- Si $\lambda = 1$, certaines de petites composantes commencent à se relier entre elles, dont la plus grande a presque sûrement une taille m de l'ordre de $N^{2/3}$ (Fig 1.15c)
- Si $\lambda > 1$, les composantes connexes continuent de s'associer pour former presque sûrement une unique **composante géante** contenant la majorité des noeuds du graphe et il n'existe plus de composantes contenant un ordre supérieur à $O(\log(N))$ (Fig 1.15d).
- Si $\lambda = N$, soit ($p = 1$), tous les couples de noeuds sont reliés, le graphe est dit **complet** (Fig 1.15e).

Cette première propriété traduit l'émergence d'une composante géante lorsque $\lambda > 1$.

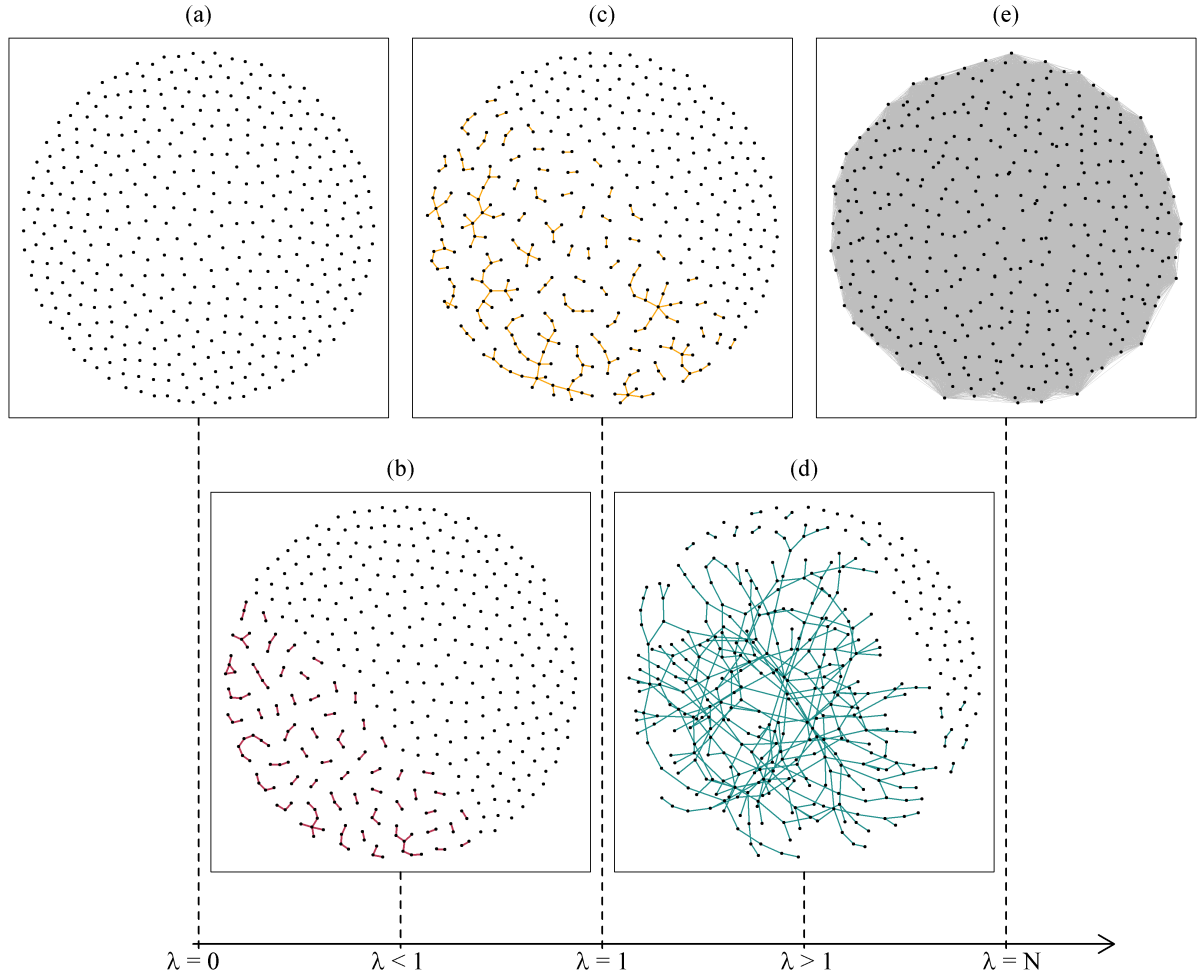


FIGURE 1.15 – Illustration des différentes phases des graphes du modèle Erdős-Rényi en fonction de $\lambda = pN$, (a) $\lambda = 0$, (b) $0 < \lambda < 1$, (c) $\lambda = 1$, (d) $1 < \lambda < N$ et (e) $\lambda = N$.

La seconde propriété remarquable du modèle d'Erdős-Rényi se situe dans la manière dont les arêtes se répartissent entre les différents noeuds. En définissant le degré d'un noeud v_i comme le nombre k de noeuds v_j adjacents à v_i selon $deg(v_i) = k$, la répartition des arêtes dans un graphe G peut être appréhendée au travers de la distribution des degrés de ses noeuds comme la fraction de noeuds de degré k , selon $P(k) = \frac{N_k}{N}$. La distribution des degrés permet de rendre compte à quel point la répartition des arêtes est homogène ou non entre tous les noeuds constituant le graphe. Cette distribution des degrés est considérée au regard du degré moyen du graphe défini comme le nombre d'arêtes divisé par le nombre de noeuds $\overline{deg}(G) = \frac{\binom{N}{2}p}{N} \approx \frac{m}{N}$.

Dans le modèle $G_{ER}(N, p)$ où la taille moyenne est de $\binom{N}{2}p$ arêtes, et considérant le tirage aléatoire et uniforme des arêtes, la probabilité qu'un noeud v pris au hasard soit de degrés k est définie par une loi binomiale [88] selon :

$$P(\text{deg}(v) = k) = \binom{N-1}{k} p^k (1-p)^{n-1-k}. \quad (1.2)$$

Ainsi, la distribution des degrés d'un graphe d'Erdős-Rényi peut être modélisée par une loi de Poisson de paramètre $\lambda = Np$ selon :

$$P(\text{deg}(v) = k) \rightarrow \frac{(Np)^k}{k!} e^{-Np}. \quad (1.3)$$

Cette convergence de la loi binomiale vers celle de Poisson est illustrée ci-après à l'aide des distributions des degrés des graphes présentés précédemment (Fig 1.15b-d) sur lesquelles sont représentées une loi de Poisson de paramètre λ propre à chaque graphe (Fig 1.16).

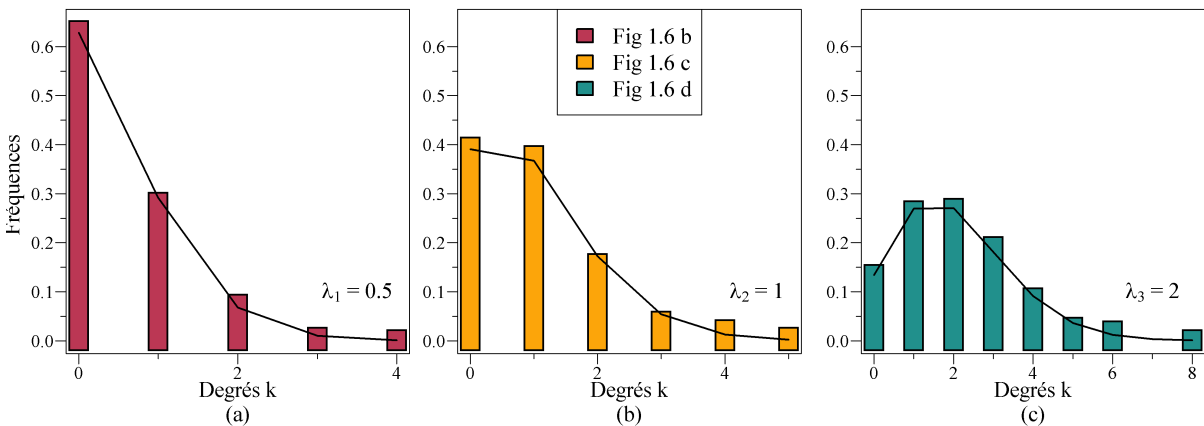


FIGURE 1.16 – Représentation de la distribution des degrés des graphes d'Erdős-Rényi (Fig 1.15b-d) et distribution d'une loi de Poisson de paramètre λ .

Ainsi, les graphes issus du modèle d'Erdős-Rényi ont été finement caractérisés aussi bien pour leur dynamique (émergence d'une composante géante) que pour leur forme statique (distribution des degrés selon un loi de Poisson). Ces caractéristiques sont le reflet d'une propriété sous-jacente à ce modèle, l'indépendance des arêtes, c'est-à-dire, que la probabilité d'apparition d'une arête entre deux noeuds, ne dépend pas de la présence ou de l'absence des autres arêtes. De ce fait, identifier qu'un graphe quelconque contient ces propriétés permet de considérer qu'un tel graphe est de type Erdős-ényi. Cette considération

doit faire écho au lecteur vis-à-vis de la problématique de ce manuscrit, l'indépendance des trajectoires des navires de pêche, devant se traduire par l'indépendance des arêtes de leur graphes de dyades.

Le modèle "*small world*" de Watts et Strogatz :

Le modèle d'Erdős-Rényi a très vite été identifié comme étant inadapté pour représenter ou analyser les phénomènes réels qui traduisent rarement, si ce n'est jamais, des relations indépendantes (il sera pour nous l'occasion de vérifier si les graphes de dyades font exceptions à cette règle).

En effet, à la fin des années 1990, après 50 ans d'analyse de graphes empiriques (notamment de la sociologie, cf. les graphes de Sampson ou de Zachary) et à la faveur du développement de nouveaux outils d'analyses qui jusqu'alors étaient essentiellement concentrés sur la distribution des degrés, des caractéristiques récurrentes reflétant l'organisation des graphes réels ont pu être identifiées dont les deux principales sont détaillées ici.

La première de ces caractéristiques est connue sous le nom de la théorie des "six degrés de séparation" introduit en 1929 par l'auteur hongrois Frigyes Karinthy évoquant l'idée que toutes les personnes sur Terre sont à moins de six connections sociales les unes des autres [89]. Cette théorie a été popularisée à partir des années 1990 sous l'appellation populaire de la "règle des six poignées de mains", ou par l'intermédiaire d'oeuvres artistiques comme récemment en 2012 par la chanson "Six Degrees of Separation" du groupe irlandais *The Script*, ou par le film du même nom de John Guare sorti en 1995 dont la bande annonce était accompagnée d'une voix off définissant cette théorie, ici traduite en français :

"Tout le monde sur la planète est séparé par seulement six autres personnes. Six degrés de séparation. Entre nous et tout le monde sur cette planète. Le président des Etats-Unis. Un gondolier à Venise... il n'y a pas seulement les grands noms. C'est n'importe qui. Un indigène dans une forêt tropicale. Un habitant de la Terre de Feu. Un eskimo. Je suis lié à tout le monde sur cette planète par une chaîne de six personnes. C'est une pensée profonde."

En terme de théorie des graphes, cette théorie fait référence au diamètre d'un graphe qui se définit par la plus grande des plus petites distances possibles qui relie 2 noeuds au sein du graphe. Ici, la notion de distance fait référence au nombre d'arêtes qu'il est nécessaire de traverser pour relier deux noeuds v_i et v_j . Le diamètre correspond alors, à la plus petite distance permettant de relier les deux noeuds v_i et v_j les plus distants du

graphe qui, selon la théorie des six degrés de séparation doit être inférieure à six. Cette théorie a été confortée aussi bien par l'observation du diamètre de graphes empiriques que par certaines expériences sociologiques dont la plus célèbre est probablement celle de Milgram [90], réduisant alors drastiquement l'idée d'un monde vaste et étendu d'où l'origine du nom que porte aujourd'hui ce phénomène, le phénomène du petit monde ("*small world phenomenon*").

Sur ce premier aspect, le modèle d'Erdős-Rényi, générant des graphes de faibles diamètres lorsque $\lambda \gg 1$, semble être adapté à la représentation des graphes réels (Fig 1.17) [87]. En effet, lorsque $\lambda < 1$, le diamètre du graphe correspond au plus grand diamètre de la multitude de composantes connexes constituant le graphes. A ce niveau d'évolution du Erdős-Rényi, certaines composantes connexes prennent la forme de longues chaînes ramifiées présentant de grands diamètres. Lorsque la transition vers la composante géante a lieu ($\lambda > 1$), ces chaînes s'associent, raccourcissant ainsi les distances entre les noeuds, responsable d'une diminution du diamètre.

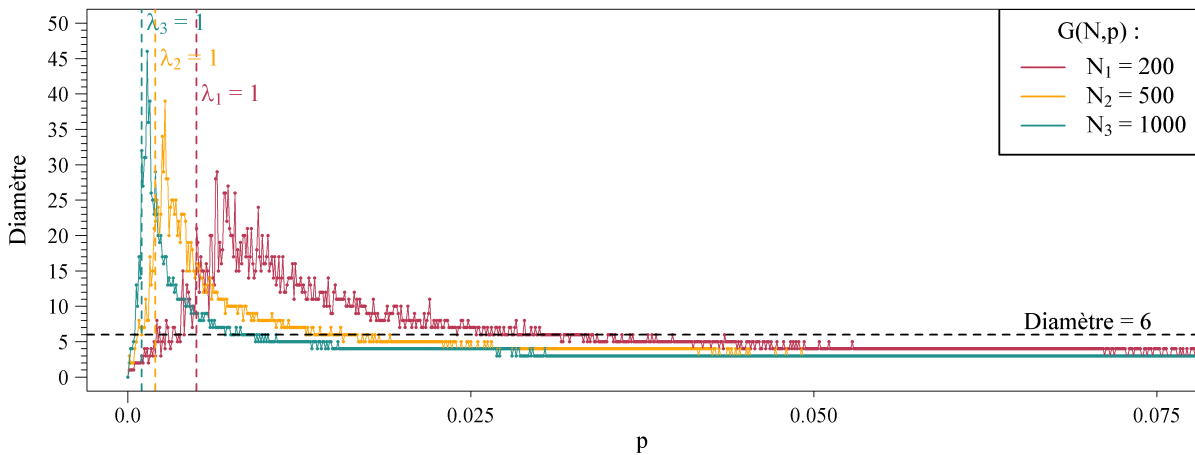


FIGURE 1.17 – Représentation de l'évolution du diamètre d'un graphe d'Erdős-Rényi d'ordre $N = \{200, 500 \text{ et } 1000\}$.

Cette première caractéristique du diamètre est relative à l'organisation globale du graphe, la seconde, en revanche, est relative à l'organisation locale du graphe, à l'échelle de groupes de quelques noeuds. En effet, les graphes réels (et encore une fois notamment ceux des réseaux sociaux) présentent la particularité d'être composés de petits groupes de noeuds fortement connectés les uns aux autres. Cette tendance des noeuds à s'agglomérer en communautés se traduit dans un graphe, par la forte abondance de structures triangulaires (trois noeuds mutuellement liés entre eux) pouvant s'expliquer du point de

vu sociologique par l'idée que "si A et B sont amis et que B et C sont amis, alors il y a de forte chance que A et C soient également amis". Cette considération sociologique se retrouve par exemple au travers des communautés décrites dans le graphe de Sampson (Fig 1.12) ou dans celui de Zachary (Fig 1.10).

En termes de théorie des graphes, ce niveau d'agglomération des noeuds sous la forme de communautés peut être appréhendé au travers d'un coefficient dit de clustering, d'agglomération ou de transitivité [91]. Ce coefficient se définit comme la probabilité que deux noeuds soient reliés entre eux en sachant qu'ils sont adjacents à un noeud commun selon :

$$C = \frac{3 \times |\text{triangles}|}{|\text{paires de voisins distincts}|} \quad (1.4)$$

C'est au niveau de cette seconde caractéristique que les graphes du modèle d'Erdős-Rényi échouent à reproduire les caractéristiques des graphes réels en présentant de faibles valeurs de coefficients de clustering.

Les premiers à s'être emparés du problème sont le sociologue Duncan Watts et le mathématicien Steven Strogatz qui proposèrent en 1998 une alternative au modèle d'Erdős-Rényi sous la forme d'un modèle de graphes aléatoires permettant de générer des graphes alliant ces deux caractéristiques, du faible diamètre associé à un fort coefficient de clustering, connu sous le nom du modèle Small World de Watts-Strogatz[91]. Dans ce modèle $G_{WS}(N, k, p)$, la génération d'un graphe est réalisée en deux temps. D'un premier temps un graphe initial circulaire d'ordre N est construit, où chaque noeud est connecté à ses k plus proches voisins(Fig 1.18a). Afin de pouvoir répartir équitablement les arêtes de chaque côté de chaque noeud soit, $k/2$ arêtes de chaque côté, ce modèle impose que k soit pair. Sous cette forme chaque noeud est alors de degré k (Fig 1.18a).

Ce graphe initial qui n'a rien d'aléatoire est un graphe particulier appartenant à la famille des graphes réguliers qui se définit par un ensemble de noeuds présentant un degré identique, ces graphes sont qualifiés de k - régulier ou k définit le degré identique de chaque noeud[92]. Sous cette forme, le graphe initial possède un très haut coefficient de clustering qui est la caractéristique ici recherchée, mais également un très grand diamètre.

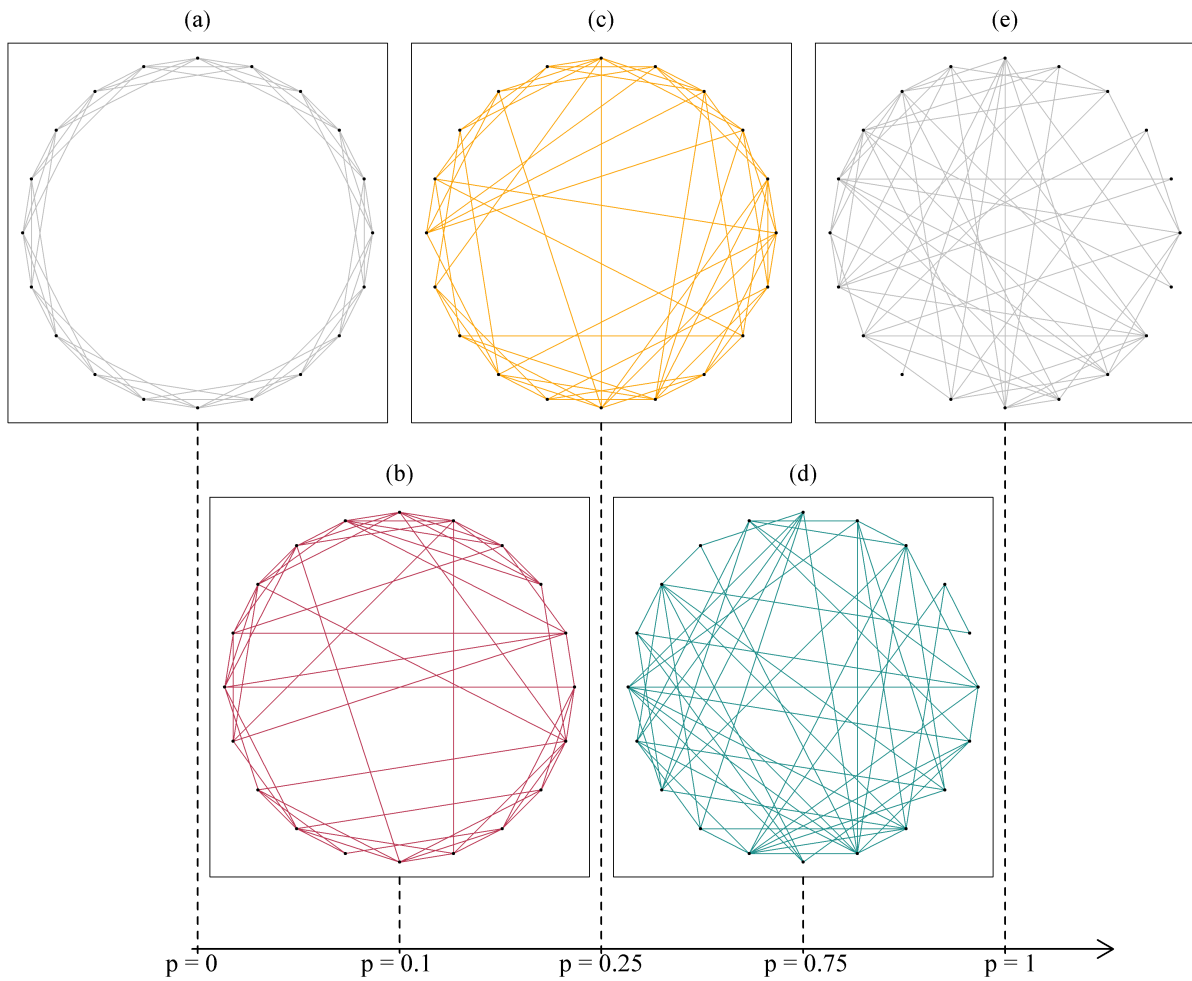


FIGURE 1.18 – Illustration de la réalisation du modèle de Watts-Strogatz $G_{WS}(20, 6, p)$ pour différents paramètres (a) $p = 0$, (b) $p = 0.1$, (c) $p = 0.25$, (d) $p = 0.75$ et (e) (a) $p = 1$

Dans un second temps, afin de réduire ce diamètre Watts et Strogatz proposent de ré-arranger aléatoirement une certaine proportion des arêtes présentes dans le graphe initial. Cette étape de ré-arrangement est contrôlée par le paramètre p définissant pour chaque arête sa probabilité p d'être reconnectée aléatoirement ou sa probabilité $1 - p$ d'être laissée en place. Autrement dit, le paramètre p contrôle le niveau de désordre à appliquer au graphe régulier initial, avec $p = 0$ (l'ordre) et $p = 1$ (le désordre). En constatant que lorsque $p = 0$ le graphe présente un coefficient de clustering maximum $C(0)$ et un diamètre maximum $L(0)$, tandis que lorsque $p = 1$, le graphe présente un coefficient de clustering minimum $C(1)$ et un diamètre minimum $L(1)$, Watts et Strogatz en ont déduit qu'il doit exister une valeur de p permettant de conserver un haut coefficient de clustering

$C(p)$ et un faible diamètre $L(p)$. En d'autres termes, il s'agit de trouver une valeur de p minimisant $\frac{L(p)}{L(0)}$ et maximisant $\frac{C(p)}{C(0)}$ (Fig 1.19).

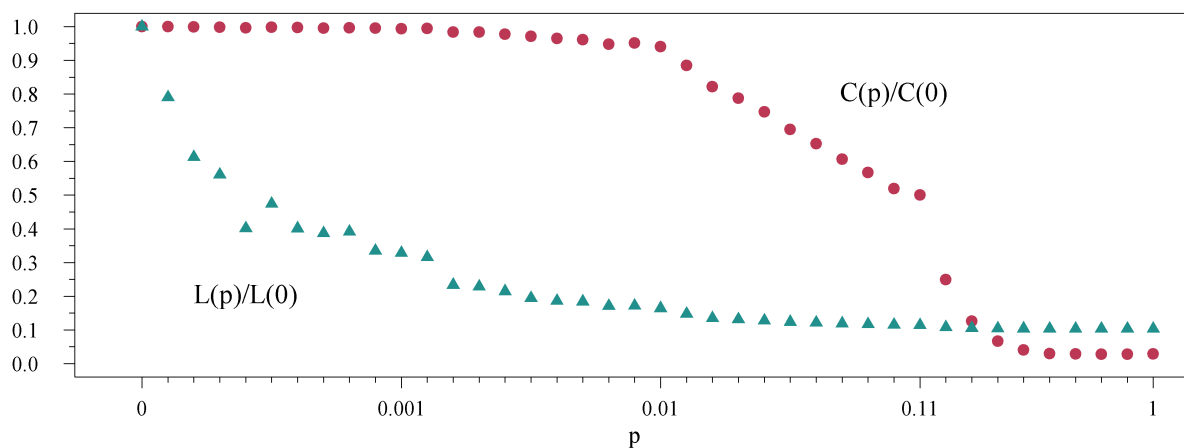


FIGURE 1.19 – Évolution du diamètre et de la transitivité d'un graphe de Watts-Strogatz $G_{WS}(1000, 20, p)$ selon le paramètre de ré-arrangement des arêtes p .

Sur la base de cette réflexion, Watts et Strogatz ont démontré qu'un faible réarrangement aléatoire des arêtes ($p \approx 0.01$) permettait de conserver un fort coefficient de clustering tout en abaissant fortement le diamètre du graphe. La cohabitation de ces deux caractéristiques est issue du réarrangement aléatoire des arêtes permettant l'émergence d'arêtes "longues portées" réduisant ainsi le diamètre du graphe.

Le modèle "*Scale Free*" de Barabasi-Albert :

De la même manière que le modèle d'Erdős-Rényi a été "écarté" du monde des graphes réels pour son inaptitude à reproduire les caractéristiques du phénomène du petit monde, le modèle de Watts-Strogatz après un engouement de faible durée a été quelque peu écarté après la découverte d'une autre particularité des graphes réels, la topologie *scale-free*.

En 1999, soit seulement une année après la parution du modèle de Watts-Strogatz, Albert-László Barabási s'aperçoit que la toute récente application internet World Wild Web présente une distribution des degrés singulière (*scale-free*) décrivant une loi de puissance selon $P(k) \sim k^{-\gamma}$ avec $\gamma = 2.1$. Il s'aperçoit que cette caractéristique n'est pas inféodée à cette seule application internet mais se retrouve dans une variété de graphes aussi distincts que les graphes de collaborations des acteurs au cinéma avec $\gamma = 2.3$ ou que celui du réseau de distribution électrique des États-Unis d'Amérique avec $\gamma = 4$ [93] (Fig 1.20). Ces observations lui permettent d'argumenter que la plupart des graphes de

grands ordre (avec $N \gg 1000$) du monde réel présente cette caractéristique qui s'avère absente des modèles de graphes d'Erdős-Rényi ou de Watts-Strogatz.

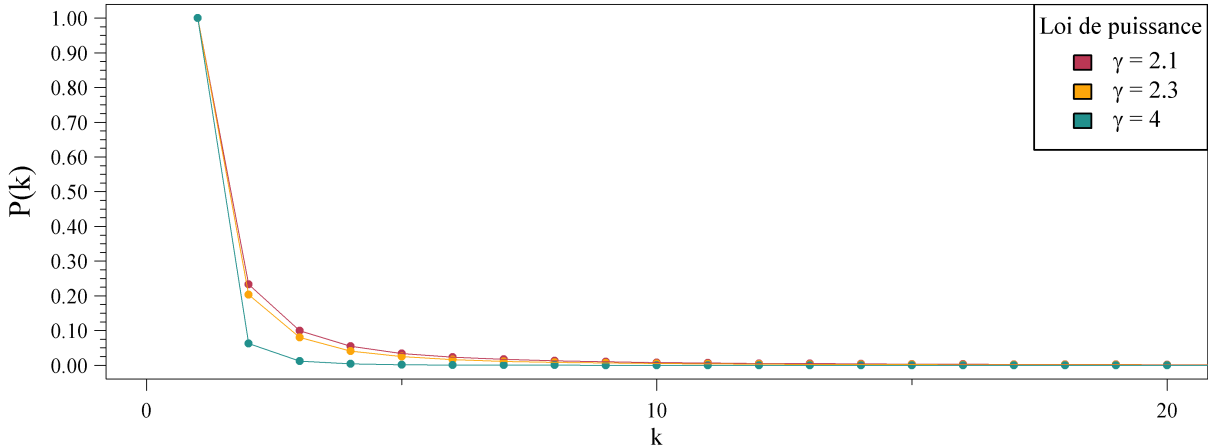


FIGURE 1.20 – Illustration des différentes loi de puissance identifiées dans les graphes réels par Albert-László Barabási [93].

En association avec Réka Albert, ils proposèrent alors un modèle de graphe aléatoire, le modèle de Barabási-Albert, incluant concomitamment deux processus sous-jacents des graphes du monde réel, jusqu'alors non pris en compte dans les modèles de graphes aléatoires, et permettant l'émergence de la propriété *scale-free*. En effet, pour Barabasi et Albert, l'émergence de la caractéristique *scale-free* de la distribution des degrés des graphes réels peut être attribuée à l'association de deux phénomènes, d'une part un phénomène de croissance (augmentation de l'ordre du graphe) du graphe, et d'autre part, un phénomène d'attachement préférentiel des noeuds entre eux.

Ainsi, dans le modèle de Barabási-Albert $G_{BA}(m_0, m)$, la génération d'un graphe se fait selon un processus itératif, où à partir d'un graphe initial d'ordre faible $N = m_0$, avec par exemple $m_0 = 2$, à chaque itération t , un nouveau noeud est ajouté emmenant avec lui m arête(s) (avec $m \leq m_0$) lui permettant de se connecter à m noeud(s) déjà présent(s) (Fig 1.21). Ce processus itératif permet d'assurer le phénomène de croissance par ajout successif de nouveaux noeuds.

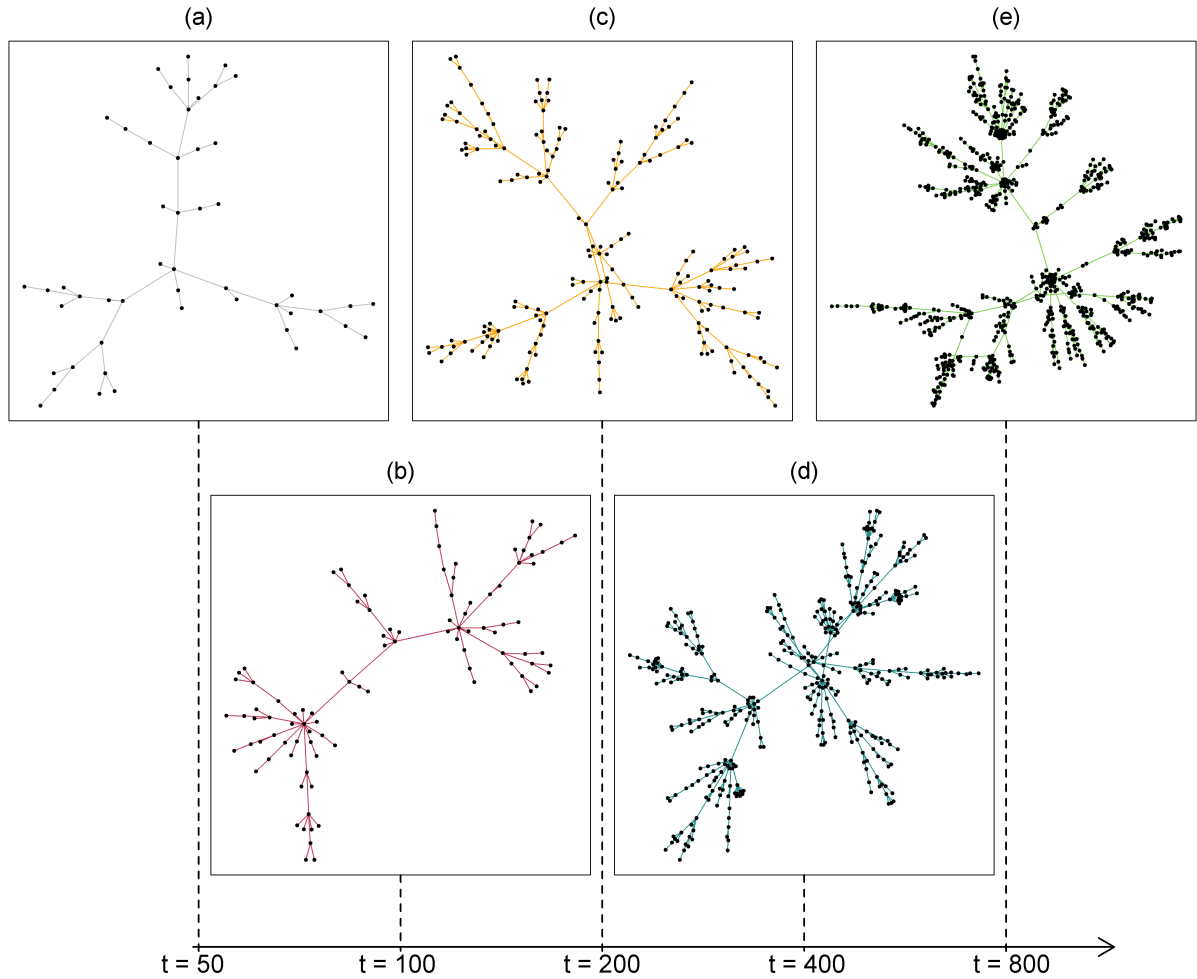


FIGURE 1.21 – Illustration de la réalisation du modèle de Barabási-Albert $G_{BA}(2, 1)$ pour différents nombre d'itérations (t) (a) $t = 50$, (b) $t = 100$, (c) $t = 200$, (d) $t = 400$ et (e) (a) $t = 800$

Le phénomène d'attachement préférentiel intervient lors de la connexion du noeud entrant avec les noeuds existants qui cette fois-ci ne se fait plus aléatoirement selon une loi uniforme, mais selon un critère d'attractivité propre à chaque noeud déjà présent. Le critère d'attractivité d'un noeud v_j déjà présent est sous la dépendance de son degré $deg(v_j) = k_j$. Ainsi, la probabilité p qu'un noeud entrant, se lie à un noeud v_j de degré k_j est définie selon $p(k_j) = \frac{k_j}{\sum_j k_j}$.

En d'autre termes, plus un noeud présente un degré élevé, plus sa probabilité d'être connecté à un noeud entrant est élevée. Ce mécanisme d'effet "boule de neige" est transcrit sous la formule du phénomène d'enrichissement des plus riches ("*rich-gets-richer phenomenon*") conduisant à l'émergence de noeuds présentant un degré "anormalement" élevé au

regard du degré moyen. En produisant un faible nombre de noeuds extrêmement connectés, appelés des "hubs", le modèle de Barabási-Albert, permet l'émergence de graphes présentant une distribution des degrés suivant une loi de puissance de paramètre $\gamma = 3$. Par la suite, le modèle de Barabási-Albert a connu certaines évolutions, notamment concernant le calcul lié à l'attachement préférentiel ici linéaire vers des formes non-linéaires et permettant de moduler le paramètre γ [94].

Et les autres modèles...

Initié par la sortie successive du modèle de Watts-Strogatz et de celui de Barabási-Albert, le début des années 2000 a été marqué par le fort intérêt de la communauté scientifique vis-à-vis des graphes, et des modèles de graphes aléatoires, s'élançant dans divers domaines scientifiques à la recherche de la propriété scale-free. Les années 2000 ont également été marquées par l'émergence et la démocratisation d'Internet et des réseaux sociaux fournissant à la communauté scientifique des bases de données permettant de produire des graphes d'ordre toujours plus élevé et plus complexes. Qu'il s'agisse de modèles originaux ou d'adaptation d'ancien, de nouveaux modèles sont apparus avec souvent la même ambition, produire des graphes mimant les caractéristiques des graphes réels dans le but d'en comprendre les phénomènes qui en sont à l'origine.

1.5 Objectif et plan de thèse :

Le travail présenté dans ce manuscrit propose d'explorer comment conceptualiser, identifier et analyser au travers de leurs trajectoires, les relations unissant un groupe d'acteurs (ici des navires de pêche) exploitant une ressource commune afin de tester leur indépendance.

Selon cette démarche, après une présentation du cas d'étude, le chapitre 2 reviendra sur l'approche présentée par Joo et al [65] permettant l'identification et l'évaluation de l'intensité des relations 2 à 2 (dyadiques) entre paires de trajectoires reconstruites à partir des données VMS (Fig 1.22-1). Ce chapitre interrogera la mise en perspective de l'ensemble de ces relations dyadiques à l'échelle d'un groupe d'acteurs au travers de leur intégration au sein de l'objet central de ce manuscrit, le graphe.

Dans un second temps, le chapitre 3 saisira à bras le corps la question de l'indépendance des trajectoires ou plutôt, celle de leurs relations dyadiques. Une première approche

reposant sur la comparaison de graphes interrogera la similarité entre les graphes empiriques de flottilles et des réalisations de modèles de graphe aléatoire et notamment le modèle d'indépendance d'Erdős-Rényi (Fig 1.22-2). Ce chapitre accordera une place importante aux méthodes de comparaison de graphes, notamment dans le cadre des petits graphes, i.e. d'ordres faibles, qui caractérisent bon nombre de situations réelles en écologie halieutique.

Enfin, le chapitre 4 sera le lieu d'un changement profond de paradigme par la remise en question de la pertinence des modèles de graphes dans l'évaluation de l'indépendance des trajectoires. L'indépendance des trajectoires ne sera plus interrogée au travers de modèles de graphes mais au travers de modèles de trajectoires. Cette modélisation des trajectoires sera permise par l'utilisation d'un nouvel outil de machine learning, les GAN ("*Generative Adversarial Network*") (Fig 1.22-3).

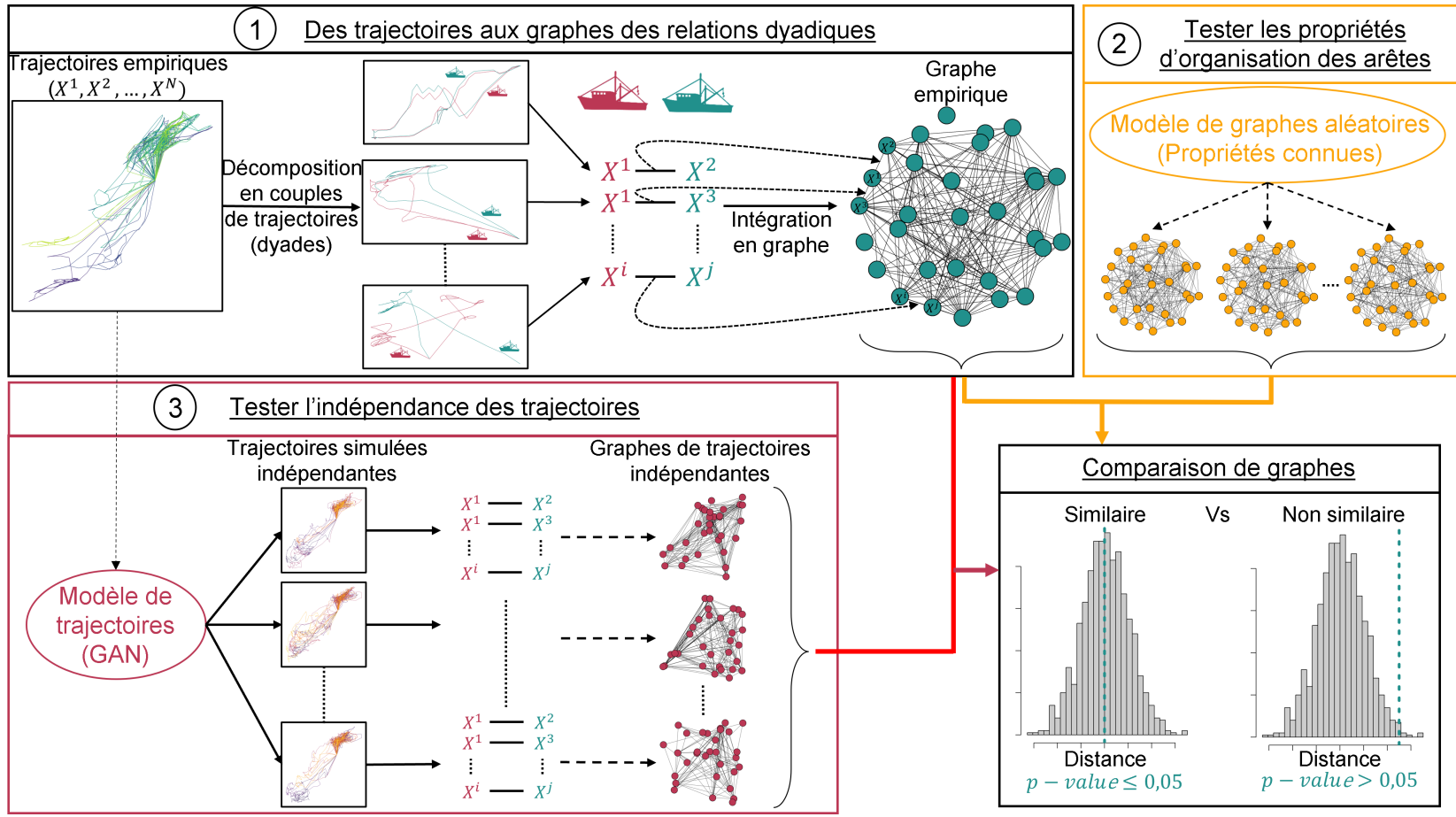


FIGURE 1.22 – Résumé graphique

DE LA TRAJECTOIRE AU GRAPHE

Dans ce deuxième chapitre, il s'agit dans un premier temps de présenter le cas d'étude traité dans ce manuscrit. Dans un second temps, au travers de son application au cas d'étude, il sera question de revenir sur l'approche développée par Joo et al[65], établissant la jonction entre l'analyse individuelle des trajectoires et l'analyse deux à deux. Cette application permettra également de présenter plus en détail les caractéristiques de la donnée VMS, ainsi que l'ensemble des traitements réalisés pour son analyse. Enfin, il s'agira d'illustrer les différentes possibilités d'intégration des relations dyadiques conduisant à la construction de différents types de graphe.

2.1 Cas d'étude et démarche

Dans cette section, nous présentons le cas d'étude choisi pour le développement de ce travail à savoir, deux flottilles chalutières françaises opérant dans la Manche entre 2012 et 2013, et d'autre part, les hypothèses de travail sous-tendant l'analyse deux à deux de leurs trajectoires.

2.1.1 La Manche, un univers de contraintes

Après avoir évoqué l'appel du large [11] et l'isolement en mer, il peut sembler ironique de définir la Manche (Fig 2.1) comme cadre d'étude. Étriquée entre la France et le Royaume-Uni, la Manche est un bras de mer s'étendant sur environ 530 km pour son plus grand axe Est-Ouest du phare de Walde (Pas-de-Calais) au phare de l'île vierge (Finistère) pour ses côtes françaises [95]. D'une forme conique, sa largeur varie entre 41 km pour son extrémité Est et 176 km pour son extrémité Ouest. Du fait de sa situation géographique de porte ouest de l'Europe, elle est le lieu d'un intense trafic maritime à l'image du rail de Ouessant représentant l'une des voies de navigation les plus empruntées au monde [96, 97].

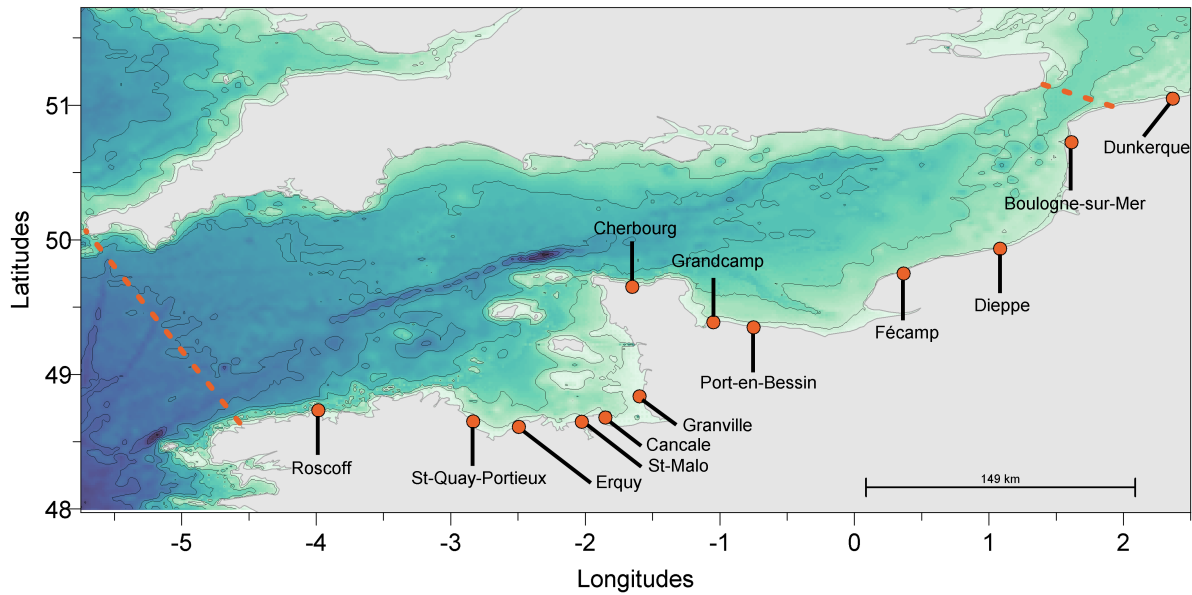


FIGURE 2.1 – Carte de la Manche et des criées françaises

La Manche est un espace de cohabitation où se côtoient navires de transport marchands, navires de plaisance et navires de pêche. Son exploitation est répartie entre plusieurs pays membres de l’UE, incluant entre autres la France, le Royaume-Uni (hors UE depuis le Brexit de 2020), l’Allemagne, l’Irlande ou encore la Belgique [98]. Concernant la France, c’est plus de 1 300 navires de pêche professionnelle [99] qui sillonnent la Manche chaque année (Table 2.1)¹.

TABLE 2.1 – Répartition des navires de pêche de la Manche selon leur taille

Années	Nombre total	Classes de taille (m)					
		(< 10)	[10; 12[[12; 18[[18; 24[[24; 40[(> 40)
2012	620	13	74	242	189	94	8
2013	652	15	99	243	190	95	10
2014	652	26	113	236	174	97	6
2015	762	58	194	237	168	96	9

La majorité de ces navires est incarnée par des navires de moins de 12m (> 70%) représentant généralement une pêche côtière (à moins de 12 milles nautiques ($\approx 22km$) des côtes) d’une grande diversité (ligneurs, tamiseurs, fileyeurs ou casayeurs) tandis que les navires restant, de plus de 12m, incarnent une pêche généralement plus hauturière

1. Les chiffres donnés dans cette table correspondent uniquement aux navires équipés du dispositif VMS, les navires de moins de 12m sont sous-représentés

reposant sur l'utilisation d'engins traînants (chalut) [100]. Pour absorber le produit de cette exploitation, 13 des 37 halles à marée (criées) que compte la France métropolitaine sont régulièrement espacées le long des côtes française de la Manche (Fig 2.1).

La combinaison d'un environnement aussi contraint que celui de la Manche et la concentration d'un grand nombre de navires de pêche, rend probable l'existence d'interactions entre les pêcheurs et l'établissement de stratégies de compétition ou de coopération entre ces derniers. Ainsi, le développement d'approches statistiques permettant d'interroger l'existence de ces comportements collectifs, en questionnant l'indépendance des navires de pêche, s'articule autour de deux flottilles de chalutiers opérant dans la Manche.

2.1.2 Flottilles d'études

En halieutique, il est courant de diviser l'ensemble d'une flotte de navires en plusieurs sous-groupes appelée flottilles, afin de constituer des groupes de navires les plus homogènes possibles au regard d'un ou de plusieurs critères [101]. Cette discrimination des navires repose généralement sur une combinaison de critères géographiques (port d'attache, région d'immatriculation, etc.) et de critères techniques (engin de pêche, taille, puissance, etc.).

Les deux flottilles considérées ici sont chacune constituées de navires de pêche de plus de 12m (équipés du dispositif VMS) opérant dans la Manche entre 2012 et 2013 et se distinguant par l'engin de pêche utilisé. La distinction des flottilles par l'engin de pêche, en conditionnant les espèces recherchées, permet d'émettre l'hypothèse d'une homogénéité des comportements de pêche entre les navires d'une même flottille alors confrontés aux mêmes contraintes. Les deux engins de pêches ont été sélectionnés sur la base de connaissance à priori, afin de constituer deux flottilles au comportement de pêche différent.

Flottille n°1 : Les chalutiers en boeufs (PTM)

La flottille des pêcheurs en boeufs, notée PTM (*Midwater/pelagic Pair Trawl*), est caractérisée par l'utilisation d'un chalut pélagique (évoluant entre deux eaux) tracté conjointement par deux navires (Fig 2.2) généralement de même dimension et de même puissance [102]. Cet engin cible des espèces pélagiques se déplaçant en banc comme le maquereau, le hareng ou le chinchard [103]. Durant l'action de pêche, les deux navires alors physiquement liés l'un à l'autre sont dans l'obligation de synchroniser leurs déplacements (direction et

vitesse). En dehors de l'action de pêche, les deux navires sont libres de leurs mouvements et peuvent choisir de prospecter l'environnement ensemble ou séparément.

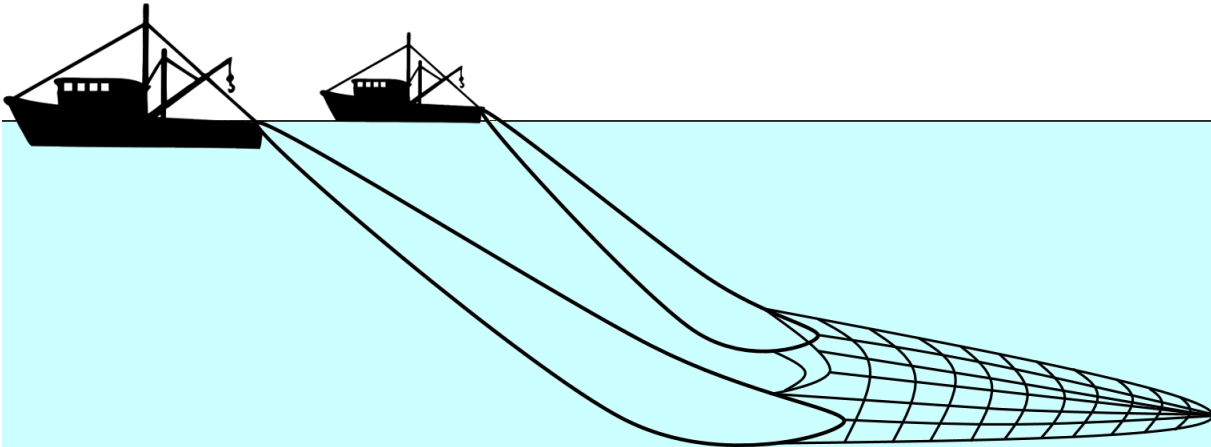


FIGURE 2.2 – Les chalutiers en boeufs ou "Pelagic Pair Trawl" (PTM)

Flottille n°2 : Les chalutiers de fond (OTB)

La flottille des chalutiers de fond, notée OTB (*Bottom Otter Trawl*) est caractérisée par l'utilisation d'un unique chalut tracté par un unique navire (Fig 2.3). A la recherche des espèces demersales (qui vivent proche du fond) comme la sole ou la lotte, le chalut est traîné sur le plancher océanique. Faisant l'objet de débats concernant sa faible sélectivité, l'érosion des fonds et la perturbation des habitats benthiques [104], le chalut de fond reste l'un des engins les plus employés de la pêche professionnelle française en Manche [100].

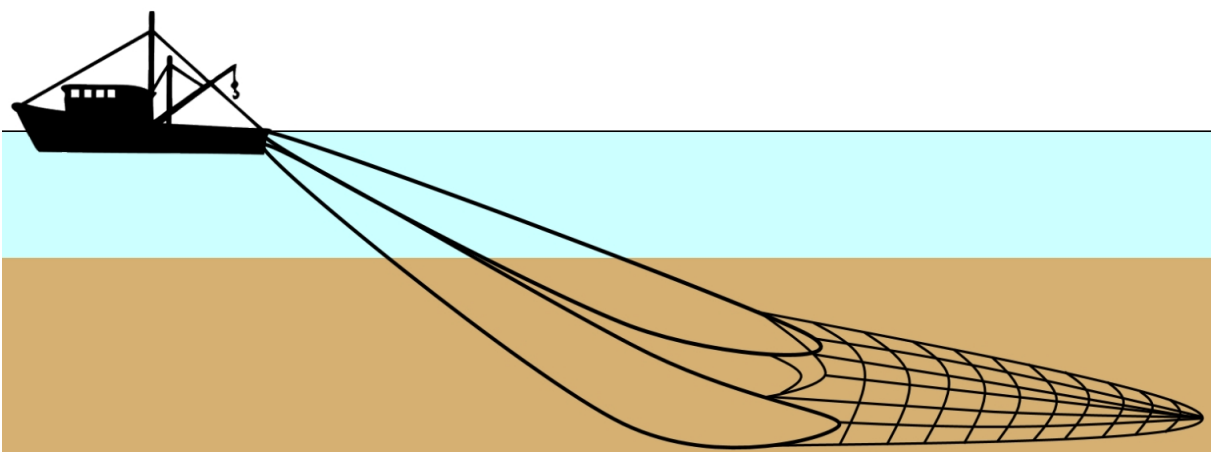


FIGURE 2.3 – Les chalutiers de fond ou "Bottom Otter Trawl" (OTB)

2.1.3 Hypothèse et démarche

La flottille des PTM, par l'utilisation simultanée d'un même engin de pêche par un couple de navires, représente une flottille de référence de l'analyse du comportement de collaboration deux à deux. Pour ce type spécifique de collaboration, il est attendu d'observer une certaine similarité dans le comportement individuel des deux membres du couple, se traduisant par une forte cohésion entre leurs trajectoires. C'est au travers de cette flottille de référence que Joo et al. [65] ont proposé le développement d'une approche reposant sur l'identification et l'analyse des relations dyadiques. Le développement de cette approche peut être traduit sur la forme d'un système d'hypothèse :

Système d'hypothèse

Hypothèse : Les relations entre pêcheurs peuvent être appréhendées au travers de la cohésion de leurs trajectoires considérées 2 à 2.

Prédiction : Les comportements de collaboration 2 à 2 se traduisent par des couples de trajectoires à forte cohésion.

Expérimentation : Identification et évaluation de la cohésion 2 à 2 des trajectoires des pêcheurs en boeufs PTM.

Observation : Voir résultats section 2.3

2.2 De la trajectoire à la dyade

Dans cette section, nous proposons dans un premier temps de présenter les caractéristiques de la donnée VMS et son utilisation dans la reconstruction des trajectoires individuelles des navires de pêche. Dans un second temps, après avoir précisé le concept de dyade, nous reviendrons sur les différentes étapes de traitement et de filtrage des trajectoires préalable à leur analyse deux à deux.

2.2.1 La donnée VMS, une succession de point dans le temps et l'espace

La donnée VMS correspond aux données de géolocalisation émises par le dispositif VMS embarqué. Cette donnée apparaît sous la forme d'une série temporelle de données spatiales mettant en relation le temps avec les coordonnées GPS de longitude et de latitude.

A l'échelle d'un navire, cette série temporelle présente une double discrétisation dans le temps (Fig 2.4).

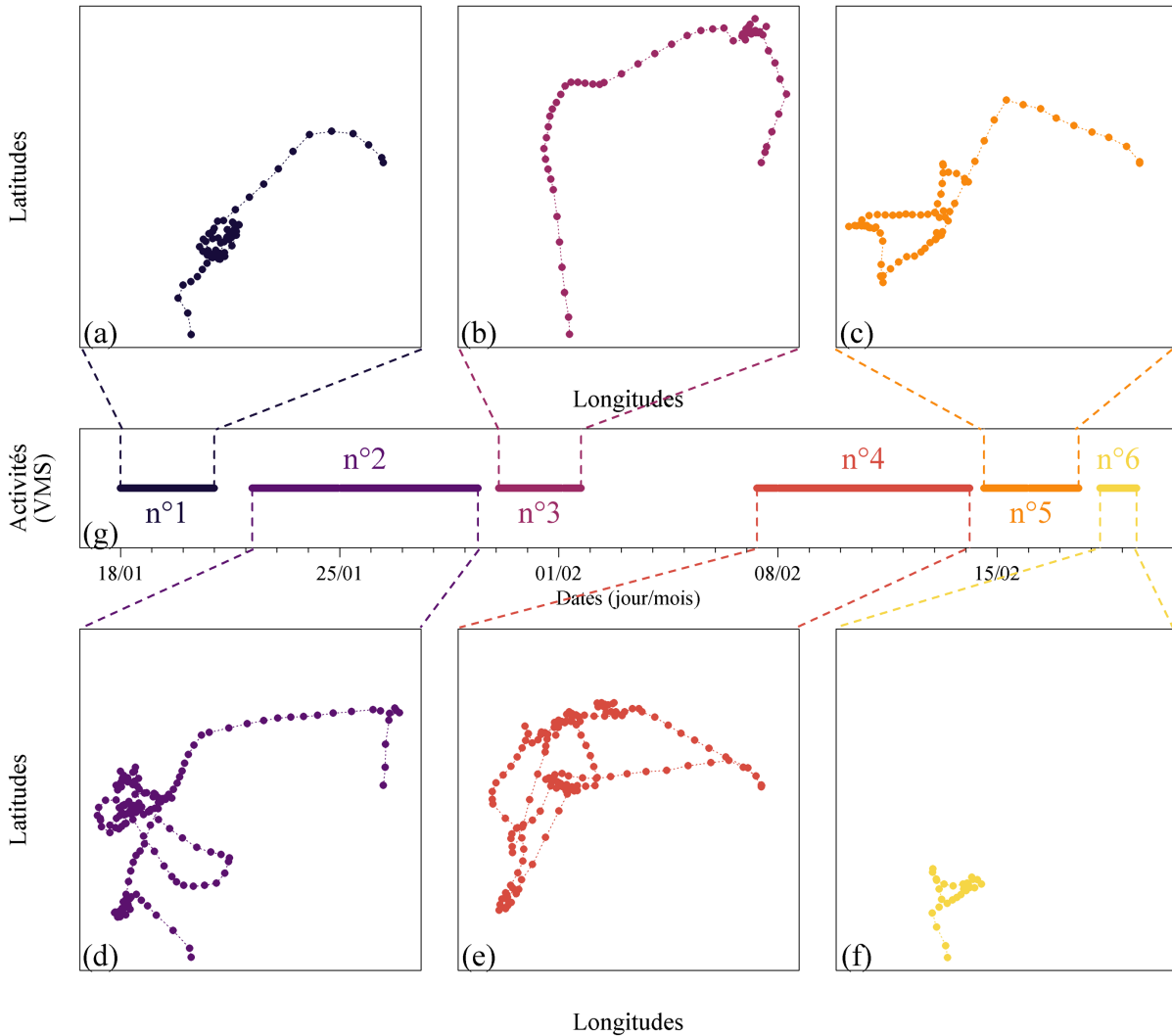


FIGURE 2.4 – Exemples de trajectoires VMS d'un même navire et sa double discrétisation.

La première discrétisation est issue de la fréquence d'enregistrement du dispositif VMS paramétré pour enregistrer un point toutes les deux heures au moins ($f \geq 0,14$ mHz). En pratique, la fréquence d'échantillonnage se situe majoritairement aux alentours d'un point toutes les heures. Ainsi, la donnée VMS prend la forme d'une série temporelle discrète \mathcal{X}_t où chaque pas de temps $t = \{t_0, t_1, \dots, t_n\}$ est associé à deux variables x_t et y_t , respectivement la longitude et la latitude (Fig 2.4a-f).

La seconde discrétisation est issue de l'activité du navire constituée d'une succes-

sion de sorties en mer entrecoupées de moments d'inactivité au port (Fig 2.4g). Il en résulte une succession de phases d'enregistrement entrecoupées de pauses. Ainsi, la série temporelle totale \mathcal{X}_t d'un navire peut être découpée en une suite de séries temporelles $\mathcal{X}_t = \{X_t^1, X_t^2, \dots, X_t^m\}$ correspondant aux m périodes d'activités du navire. En termes halieutiques, ces périodes d'activités sont appelées marées.

2.2.2 L'activité de pêche, une succession de marées

Formellement une marée est définie comme la période d'activité d'un navire de pêche qui s'écoule depuis l'appareillage jusqu'au retour au port. Une marée comprend alors à la fois des phases non productives (route, recherche, traitement des captures, avarie de matériel, mise à la cape par gros temps, etc.) et des phases effectives de pêche (engin déployé). Dans ce travail, la notion de trajectoire est liée à celle de la marée et se définit comme la série temporelle de géolocalisation (X_t) d'un navire au cours de cette dernière. A l'échelle d'un navire, il existe donc autant de trajectoires, qu'il existe de marées.

De plus, une marée correspond ici à l'utilisation d'un engin de pêche spécifique qui, pour un même navire, peut dans certain cas, varier d'une marée à l'autre traduisant d'un changement de stratégie ou d'espèce cible généralement conditionné par le changement de saison. La constitution des flottilles apparaît alors comme plus restrictive que le simple regroupement d'un ensemble de navires, mais plutôt comme le regroupement d'un ensemble de marées caractérisées par l'utilisation d'un même engin de pêche. Un même navire, peut être référencé dans plusieurs flottilles selon les engins de pêche qu'il utilise.

Selon ces considérations, le regroupement des marées en deux flottilles PTM et OTB, conduit à ce que d'un côté la flottille PTM soit clairement représentée par des navires de 18 à 24m représentant à eux seuls plus de 80% des marées de cette flottille (Table 2.2). De l'autre, la flottille OTB est constituée de navires partagés en trois classes de taille s'étendant de 12 à 40m, traduisant des différences significatives de leurs caractéristiques techniques.

Ainsi, le regroupement des marées en flottille n'exclut pas l'existence d'une certaine variabilité spatiale et temporelle des caractéristiques de leurs marées. Par exemple, la durée de ces dernières peut présenter une forte variabilité en s'étendant de quelques heures à plusieurs jours (Fig 2.5a). L'effet de la taille des navires sur la durée des marées est très nettement visible, notamment concernant les navires OTB dont la durée moyenne des marées augmente avec la taille des navires (Fig 2.5b-d). Toutefois au-delà de ces caracté-

TABLE 2.2 – Répartition en classes de taille (m) des navires et des marées PTM et OTB des années 2012 et 2013.

Engin	Effectif	Classes de taille (m)					Total
		[10; 12[[12; 18[[18; 24[[24; 40[(> 40)	
PTM	Navires	1 (1.9%)	11 (21.2%)	37 (71.2%)	3 (5.8%)	0 (0%)	52
	Marées	1 (0.1%)	161 (13.6%)	987 (83.5%)	33 (2.8%)	0 (0%)	1182
OTB	Navires	0 (0%)	135 (52.3%)	72 (27.9%)	50 (19.4%)	1 (0.4%)	258
	Marées	0 (0%)	9022 (48.4%)	6119 (32.8%)	3494 (18.7%)	2 (0%)	18637

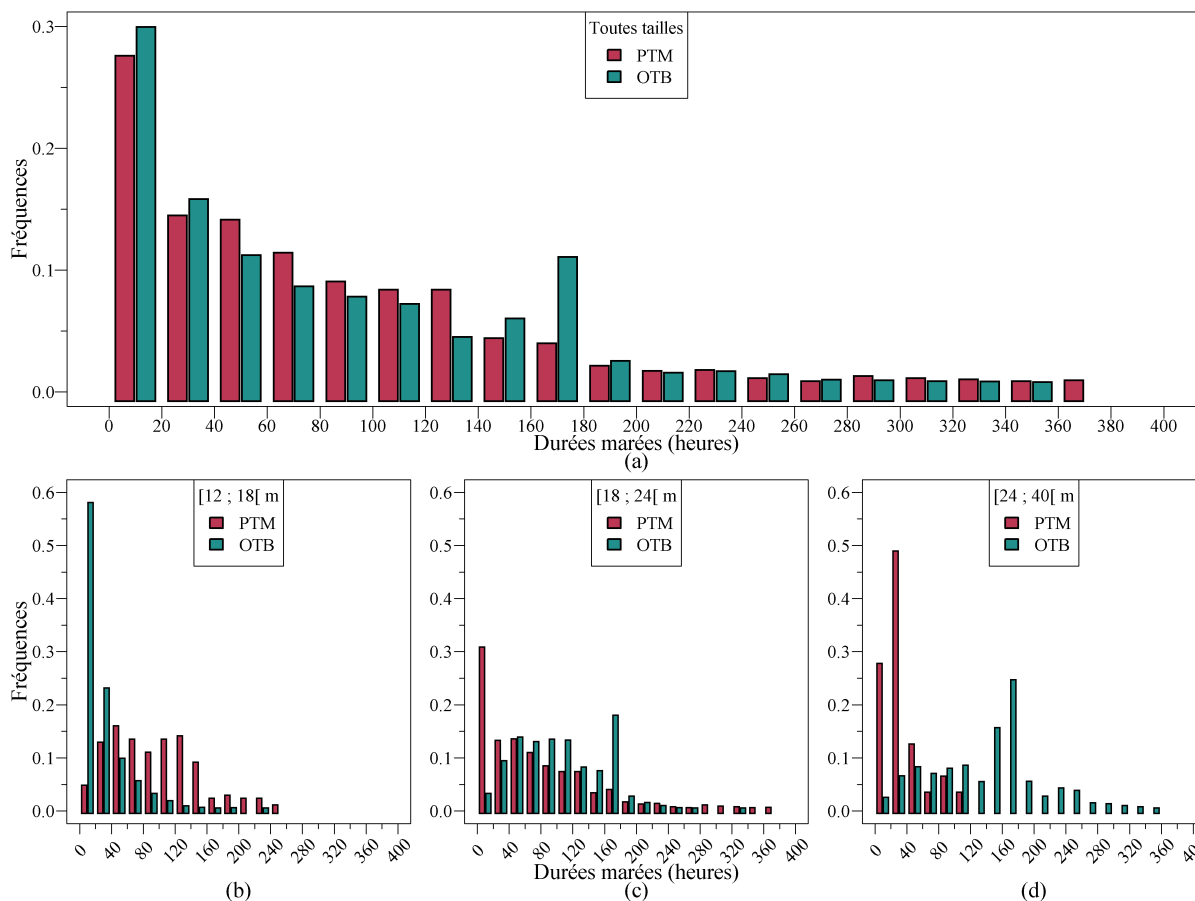


FIGURE 2.5 – Histogrammes des durées marées PTM et OTB selon la taille des navires .

ristiques techniques imposant une autonomie maximale indépassable, la durée des marées (d'un même navire ou non) est conditionnée par d'autres facteurs tels que l'efficacité de la

pêche (temps nécessaire pour remplir les cales, les quottas ou les objectifs de capture), de la stratégie des pêcheurs, de l'évolution des conditions météorologiques ou d'éventuelles avaries.

De même une variabilité spatiale perdue, tandis que certains navires sont inféodés à un unique port de rattachement, les restreignant à exploiter les mêmes zones spatiales dont l'étendue dépend ici également des caractéristiques techniques du navire, d'autres présentent la particularité de se déplacer d'un port à l'autre (Fig 2.4a,b,d) leur permettant potentiellement d'élargir leur zone d'exploitation.

2.2.3 La cohésion deux à deux, la brique fondamentale

S'intéresser au comportement deux à deux permet de décomposer un ensemble de trajectoires (Fig 2.6) en une somme de couples de trajectoires, réduisant ainsi la notion de comportement collectif à son niveau d'expression le plus bas. Sur cette projection, seulement quinze jours d'enregistrement VMS sont suffisants pour recouvrir la quasi totalité de la Manche de trajectoires, révélant ainsi l'intense activité de pêche qui s'y établit.

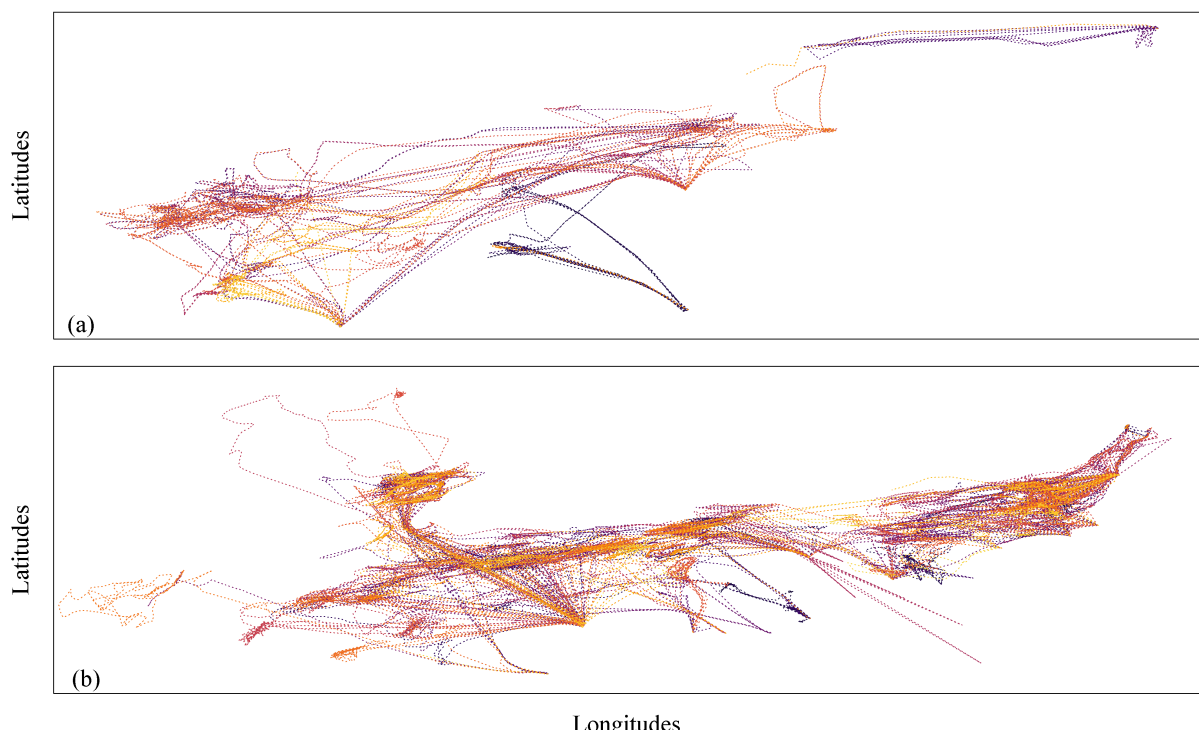


FIGURE 2.6 – Illustration de quinze jours d'enregistrement VMS des trajectoires des flottilles (a) PTM et (b) OTB.

La notion dyade :

La configuration particulière des PTM en pêcheurs en boeufs conduit à la formation de couples de trajectoires au comportement singulier, imposé par deux contraintes liées à l'utilisation d'un même engin de pêche (Fig 2.2). D'une part, le partage d'un même espace temporel, il s'agit d'être ensemble au même moment en mer et, d'autre part, le partage d'un même espace spatial, il s'agit d'être ensemble au même endroit afin de manipuler le même engin de pêche.

C'est sur la base de ces deux contraintes, temporelle et spatiale, que Joo et al [65] ont défini un objet élémentaire permettant d'appréhender les comportements 2 à 2, la dyade. Une dyade se définit comme un couple de trajectoires partageant le même espace temporel et spatial, c'est-à-dire issues deux navires simultanément en mer ayant été au moins une fois simultanément au "même endroit". Ainsi, une dyade provenant d'un couple de navires PTM travaillant ensemble, est constituée de deux trajectoires presque identiques (Fig 1.7a). Dans ce cas précis, où les deux navires sont partis et revenus au même moment du même port (ici Boulogne-sur-Mer), les trajectoires constituant la dyade correspondent à l'entièreté des marées des deux navires, mais il n'en est pas toujours ainsi. En effet, une dyade peut être constituée uniquement des segments des trajectoires où les navires sont simultanément en mer (Fig 1.7b). Dans cet autre cas, un des deux navires était déjà en mer lorsque le second a appareillé aboutissant à une dyade ne couvrant qu'une fraction de leurs marées respectives. Dans ce second exemple, la dyade est constituée uniquement parce que les deux navires se sont retrouvés simultanément au même endroit au terme d'une route de croisement.

A ce stade, l'identification des dyades au sein d'un ensemble de trajectoires permet seulement d'identifier les couples de navires présents simultanément en mer et ayant potentiellement pu entretenir une relation, par exemple collaborative. La dyade apparaît alors comme le niveau 0 de la compréhension du comportement 2 à 2, elle est nécessaire mais insuffisante.

L'identification des dyades est alors complétée par leur description au travers d'un ensemble de statistiques descriptives [105] permettant d'interroger la cohésion entre les trajectoires constituant une dyade (Fig 2.7). Ainsi, pour deux segments de trajectoires X_t^A et X_t^B de deux navires A et B formant une dyade de durée T (nombre de pas de temps), trois métriques ont été choisies afin d'interroger chacune un aspect différent de la cohésion deux à deux.

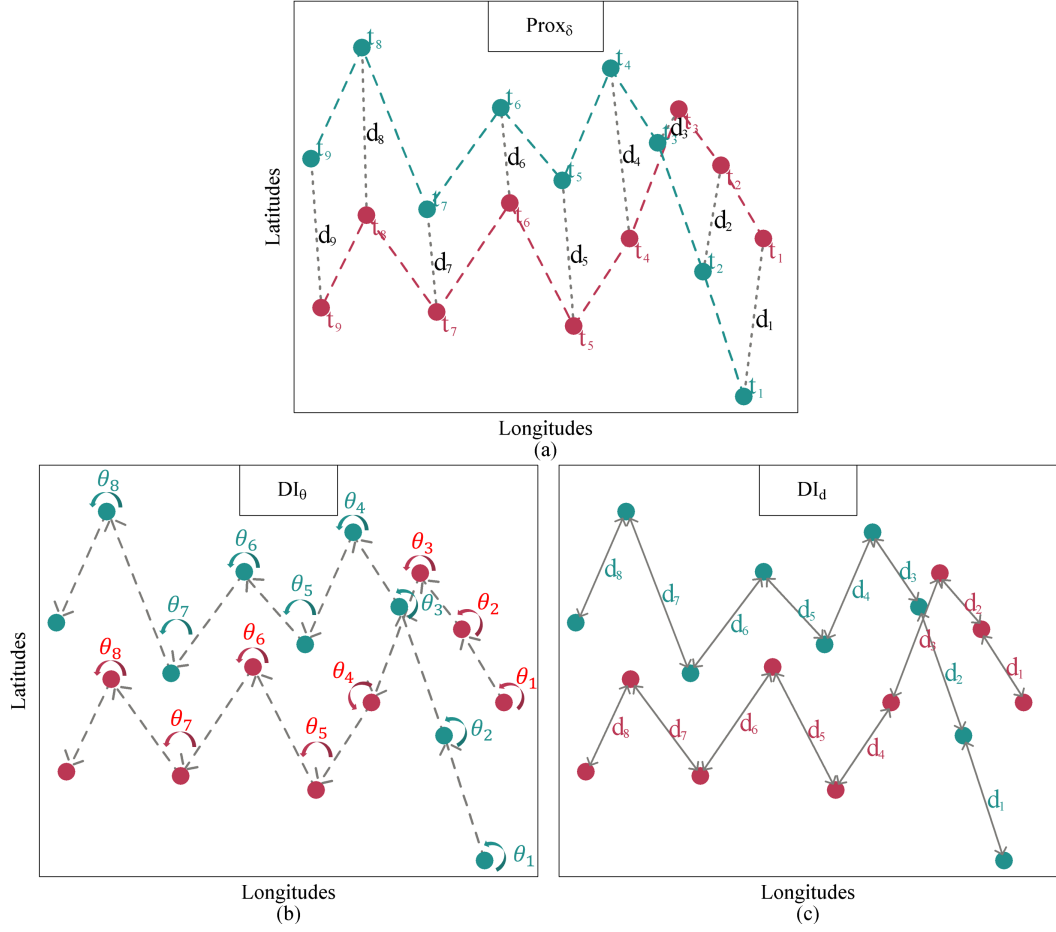


FIGURE 2.7 – Représentation des trois métriques descriptives de la cohésion des trajectoires formant une dyade (a) $Prox_\delta$, (b) DI_θ , (c) DI_d .

Statistique de proximité ($Prox_\delta$) :

L'indice $Prox_\delta$ [106] (Fig 2.7a) renseigne pour un couple de trajectoires, leur niveau de proximité physique moyen selon,

$$Prox_\delta = \frac{1}{T} \sum_{t=1}^T \left(\{dist(X_t^A, X_t^B) < \delta\} \right), \quad (2.1)$$

soit la proportion de pas de temps où les deux navires A et B sont simultanément à une distance inférieure à un seuil δ . Ainsi, l'indice $Prox_\delta$ varie selon $0 \leq Prox_\delta \leq 1$ avec $Prox_\delta = 0$, les deux navires n'ont jamais été simultanément à une distance inférieure à δ , et $Prox_\delta = 1$, les deux navires ont toujours été simultanément à une distance inférieure à δ . La constitution d'une dyade entre deux trajectoires telle que décrite précédemment est

conditionnée par un critère de proximité ($\delta = 5km$), en conséquence, toutes les dyades présentent une métrique de proximité $Prox_\delta > 0$.

Statistique de synchronie angulaire (DI_θ) :

L'indice DI_θ [107] (Fig 2.7b) renseigne pour un couple de trajectoires, leur propension à se déplacer simultanément dans la même direction selon,

$$DI_\theta = \frac{\sum_{t=1}^{T-1} \cos(\theta_t^A, \theta_t^B)}{T-1}, \quad (2.2)$$

avec θ_t^A et θ_t^B les angles décrits par les trajectoires entre les temps t et $t+1$ par rapport au plan horizontal (azimut). L'indice DI_θ renseigne en moyenne, la différence d'angle entre les deux trajectoires et varie selon $-1 \leq DI_\theta \leq 1$ avec $DI_\theta = -1$, les deux navires vont toujours dans des directions opposées et $DI_\theta = 1$, les deux navires vont toujours dans la même direction.

statistique de synchronie de déplacement (DI_D) :

L'indice DI_D [107] (Fig 2.7c) renseigne pour un couple de trajectoires, leur propension à parcourir simultanément la même distance (\approx à se déplacer à la même vitesse) selon,

$$DI_D = \frac{\sum_{t=1}^{T-1} \left(1 - \left(\frac{|d_{t,t+1}^A - d_{t,t+1}^B|}{d_{t,t+1}^A + d_{t,t+1}^B} \right)^\beta \right)}{T-1}, \quad (2.3)$$

avec $d_{t,t+1}^A$ et $d_{t,t+1}^B$ la distance parcourue entre les temps t et $t+1$. L'indice DI_D renseigne, la différence des distances parcourues entre les deux trajectoires et varie selon $0 \leq DI_D \leq 1$ avec $DI_D = 0$ les deux trajectoires ont parcouru des distances très différentes et $DI_D = 1$, les deux trajectoires ont parcouru les mêmes distances au même moment. Le paramètre $\beta = 1$ par défaut, et utilisé ici, module la sensibilité à la différence de mouvement.

Ainsi, l'association de ces trois métriques permet d'évaluer quantitativement la cohésion entre les deux segments de trajectoires formant une dyade (Table 2.3). Concernant les deux exemples de dyades précédents (Fig 1.7), avec un seuil de distance nécessaire à l'indice de proximité fixé à $\delta = 5km$, la dyade formée par les navires PTM en pleine collaboration (Fig 1.7a) présente des métriques de très forte valeur traduisant la forte cohésion liant leur deux trajectoires. A l'inverse, la dyade formée par les navires OTB, sans

comportement spécifique a priori et s'étant seulement croisés en mer (Fig 1.7b), présente de métriques notamment de proximité et de synchronie angulaire de faible valeur, traduisant une faible cohésion entre leur trajectoires et rendent compte du simple croisement des deux navires en mer.

TABLE 2.3 – Exemple des métriques de deux dyades (Fig 1.7)

Métriques	Dyade n°1 (Fig 1.7a)	Dyade n°2 (Fig 1.7b)
$Prox_{\delta=5km}$	0.821	0.010
DI_{θ}	0.980	0.480
DI_D	0.947	0.804

Si le comportement liant deux trajectoires doit être appréhendé au travers de leur cohésion, traduite sous la forme de métriques descriptives, il est nécessaire de s'assurer que ces dernières soient les plus pertinentes et précises possible. Or cette pertinence et cette précision sont sous la dépendance de deux facteurs, d'une part la durée de la dyade, conditionnant le nombre de points sur lesquels les métriques sont évaluées, et d'autre part, le format de la donnée VMS d'entrée.

2.2.4 Synchronisation et filtrage de la donnée VMS

Les métriques descriptives présentées précédemment reposent sur une analyse terme à terme (pas de temps à pas de temps) des caractéristiques simultanées (proximités, angles et distances) entre les trajectoires formant une dyade. Or, la procédure d'enregistrement de la données VMS a été mise en place dans le cadre d'un suivi individuel des navires de pêche en excluant l'éventualité du détournement de son utilisation dans le cadre d'un suivi collectif, conjoint et simultané de ces mêmes navires, aboutissant à des séries temporelles brutes pouvant présenter une asynchronie.

La donnée VMS, une donnée asynchrone :

Malgré l'absence d'une procédure d'enregistrement standardisée de la donnée VMS ayant pour seul pré-requis "*l'enregistrement d'au moins un point toutes les deux heures*", la majorité des fréquences d'échantillonnages des points VMS restent similaires entre les différents navires (Fig 2.8) oscillant autour d'un point toutes les heures environ. Il est tout de même nécessaire de corriger ces différences, même faibles, afin d'uniformiser les fréquences d'échantillonnages.

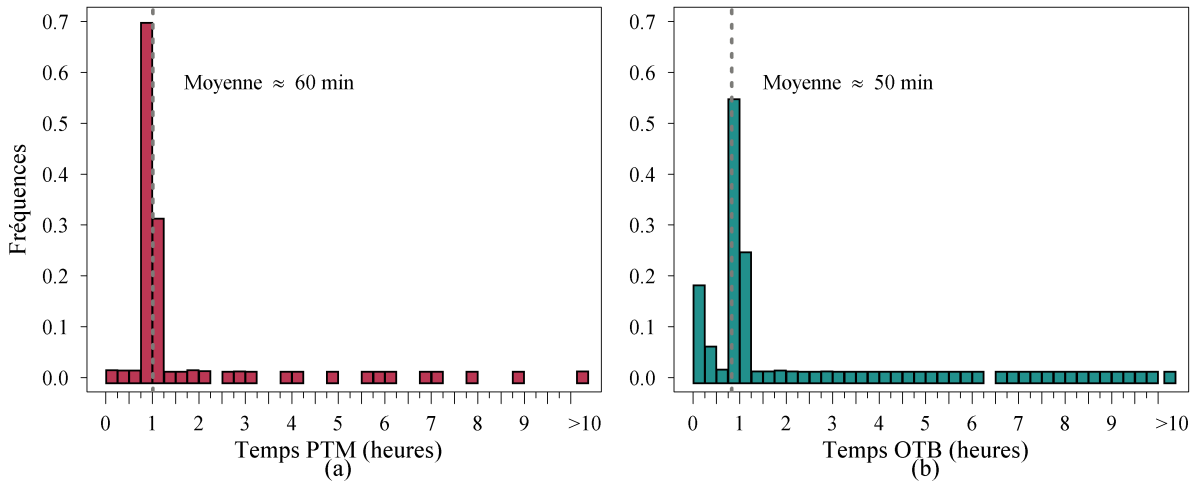


FIGURE 2.8 – Histogrammes des pas de temps entre deux points VMS (a) PTM (b) OTB.

En général, la fréquence d'échantillonnage ne varie pas au cours d'une même marée alors décrite par des points VMS régulièrement espacés. Cependant, deux phénomènes peuvent conduire à une asynchronie entre deux trajectoires (Fig 2.9).

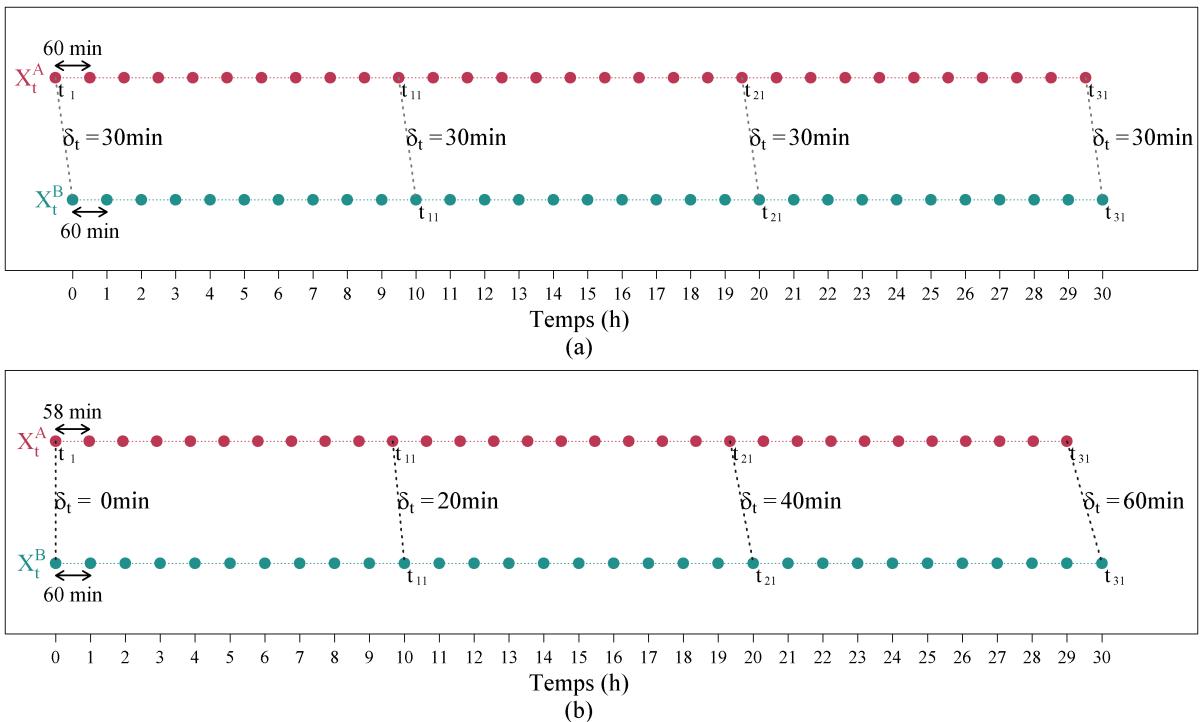


FIGURE 2.9 – Illustration de l'asynchronie des données VMS (a) fréquences d'échantillonnages identiques et (b) fréquences d'échantillonnages différentes.

D'une part, considérons deux trajectoires X_t^A et X_t^B présentant une fréquence d'échantillonnage f_A et f_B identique, par exemple, à raison d'un point toutes les heures (Fig 2.9a). Dans ce cas, une asynchronie temporelle entre les deux trajectoires peut exister si les "moments" de l'échantillonnage diffèrent. En d'autres termes deux dispositifs VMS échantillonnant à la même fréquence peuvent ne pas être en phase.

D'autre part, considérons deux trajectoires X_t^A et X_t^B présentant une fréquence d'échantillonnage f_A et f_B différente, par exemple à raison d'un point toutes les 58 minutes pour l'une et toutes les 60 minutes pour l'autre (Fig 2.9b). Dans ce second cas, avec un point initial (t_1) échantillonné à la même heure pour les deux trajectoires, la différence de 2 minutes dans le pas de temps d'échantillonnage s'accumule de proche en proche pour atteindre au bout de 30 heures, une différence d'une heure entre les deux derniers points.

Ces deux types d'asynchronie, de phase et/ou de fréquence nécessitent d'être préalablement corrigés avant l'identification des dyades et l'application des métriques descriptives de la cohésion simultanée entre les couples de trajectoires. En effet, en absence de synchronisation, par exemple l'indice de proximité $Prox_\delta$ calculé entre deux trajectoires asynchrones ne refléterait pas une proximité instantanée mais une proximité différée dans le temps, entre la position instantanée d'un des deux navires et celle qu'occupait ou qu'occupera le second un moment plus tôt ou plus tard. Le raisonnement est le même pour les deux autres métriques DI_D et DI_θ .

Enfin, dans les cas les plus critiques même si minoritaires, il peut même être observé sans raisons connues (dysfonctionnement du dispositif VMS ou manipulation humaine), une variation de la fréquence d'échantillonnage au cours d'une même marée, faisant varier le pas de temps d'une dizaine de minute à plus d'une heure (Fig 2.10c).

Resynchronisation de la donnée VMS :

La resynchronisation est une étape courante de transformation de la donnée brute en analyse de trajectoire. Cette opération de transformation est ici réalisée au travers d'une interpolation linéaire [108] de chaque trajectoire permettant de rendre artificiellement toutes les données VMS synchrones dans le temps (Fig 2.10). L'interpolation linéaire consiste à ré-échantillonner les séries temporelles brutes en estimant la position de nouveaux points à des temps prédéterminés. Ainsi, soit X_t la série temporelle discrète d'une trajectoire quelconque échantillonnée au temps $t = \{t_1, t_2, \dots, t_n\}$, la méthode repose sur l'hypothèse que la trajectoire parcourue entre les temps t et $t + 1$ peut être approchée

par une fonction affine. En posant $f(t) = x_t$ et $f(t + 1) = x_{t+1}$ la longitude aux temps t et $t + 1$, alors pour tout temps t' tel que $t \leq t' \leq t + 1$ il existe un point de la fonction linéaire correspondant à $f(t') = x_{t'}$ tel que :

$$f(t') = \frac{(t + 1) - t'}{(t + 1) - t}x_t + \frac{t' - t}{(t + 1) - t}x_{t+1} \quad (2.4)$$

Il en va de de même concernant les latitudes y_t .

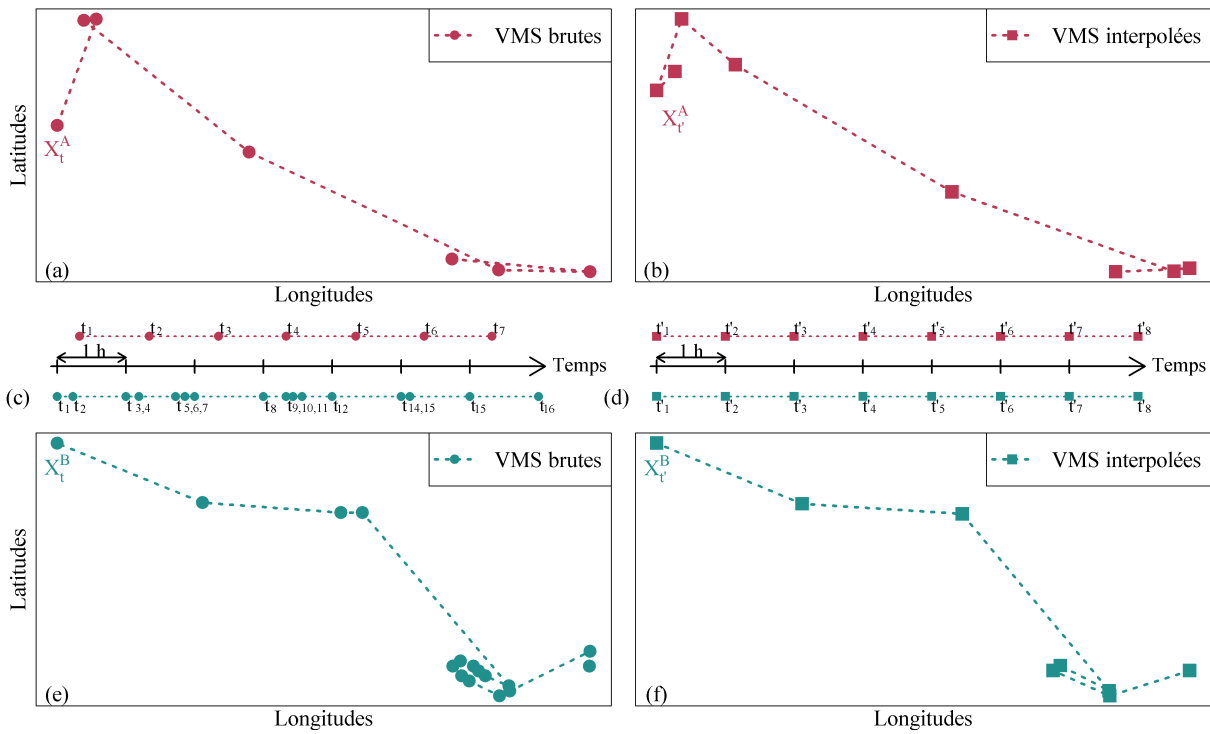


FIGURE 2.10 – Illustration de la resynchronisation de deux trajectoires par l'interpolation linéaire de deux trajectoires VMS échantillonnées à des fréquences différentes.

En définissant $\delta'_t = (t' + 1) - t' = 1h$, l'application d'une interpolation linéaire à l'ensemble des séries temporelles permet d'uniformiser leurs fréquences d'échantillonnages et ainsi, obtenir un pas de temps régulier entre deux points successifs. De plus, en définissant les temps d'interpolation comme des heures pleines $t' \in \{00h00, 01h00, \dots, 23h00\}$, l'interpolation linéaire permet de mettre en phase la nouvelle fréquence d'échantillonnage (Fig 2.10).

Limites à la resynchronisation, les données manquantes :

Bien que l'interpolation linéaire permette de réaliser cette étape nécessaire de resynchronisation et d'homogénéisation des données, elle reste une méthode d'estimation dont le résultat est une version transformée de la donnée d'entrée. D'un ensemble de positions réelles, les trajectoires des marées sont transformées en un ensemble de positions estimées. La précision de l'estimation est dépendante de la paramétrisation de l'interpolation et notamment du différentiel entre les fréquences d'échantillonnages d'entrée et de sortie. En effet, il est toujours préférable de produire par interpolation, qu'elle soit linéaire ou non, une version sous-échantillonnée de la version d'entrée plutôt qu'une version sur-échantillonnée. En d'autres termes, produire une série temporelle présentant un point toutes les heures à partir d'une série temporelle présentant un point toutes les dix minutes est préférable à l'opération inverse.

Ici, la fréquence d'échantillonnage de l'interpolation linéaire permet de produire des séries temporelles présentant un pas de temps proche (1h) de celui observé en moyenne dans les données VMS brutes (1h pour les PTM (Fig 2.8a), et 50 minutes pour les OTB (Fig 2.8b)). En d'autres termes, dans la plupart des cas, l'interpolation linéaire consiste à réajuster la position des points dans le temps sans ajout ni suppression.

Toutefois, certaines trajectoires présentent des données manquantes, appelées ici "gaps", qui s'établissent sur des temps plus ou moins longs ($> 2h$) (Fig 2.8). Ces gaps pourraient être comblés par l'ajout de points supplémentaires lors de l'interpolation mais consisteraient à un cas de sur-échantillonnage. L'origine de ces gaps n'est pas totalement établie ou pour certains il peut s'agir d'un simple dysfonctionnement du dispositif VMS tandis que pour d'autres, une manipulation humaine, volontaire ou non, ne peut pas être exclue. Néanmoins, pour certains d'entre eux, il existe une origine clairement identifiée, issue du processus d'acquisition des données VMS.

La donnée VMS est une donnée dite sensible et confidentielle devant subir avant utilisation un pré-traitement visant à les anonymiser et n'est accessible que sur demande. Au cours de cette demande, une zone géographique est spécifiée (ici la Manche) et la donnée VMS filtrée en fonction de la zone requise est transmise. Or, selon la requête les données VMS en dehors de la zone spécifiée ne sont pas transmises, et ce, même si elles font partie d'une trajectoire s'établissant en partie sur la zone spécifiée. Ainsi, un certain nombre de marées (environ 25%) se trouvent amputées d'une partie de leur trajectoire située en

dehors de la zone d'étude (Fig 2.11)².

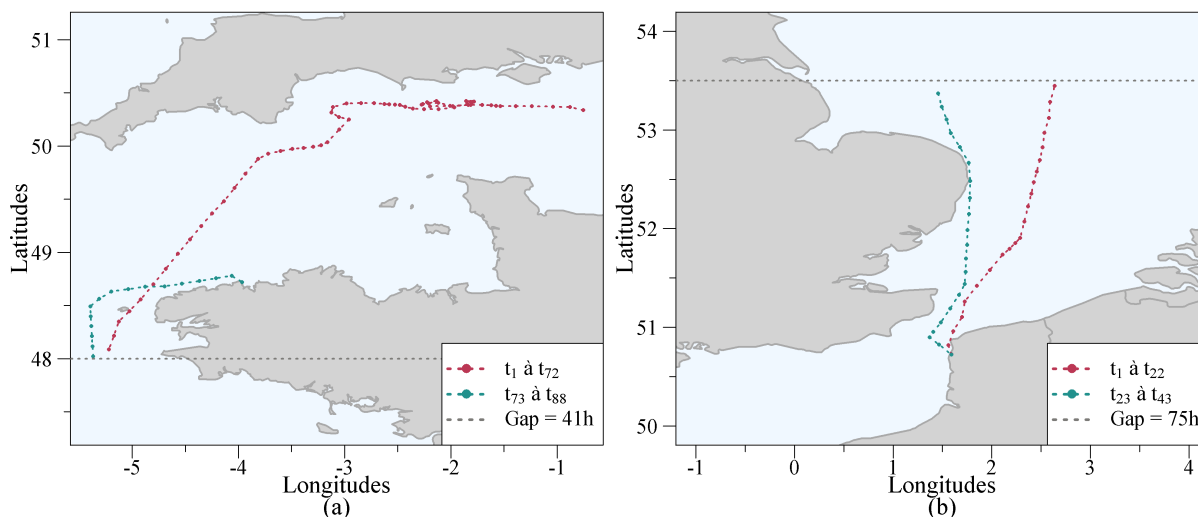


FIGURE 2.11 – Illustration de deux trajectoires incomplètes (gap) par leur sortie de la zone d'étude (a) par la frontière sud-ouest de la manche (b) par la frontière nord-est de la Manche.

Afin de limiter le nombre de marées incomplètes, un premier filtrage des trajectoires est réalisé en amont de tout traitement, visant à ne conserver que les trajectoires ayant à la fois comme ports de départ et de retour des ports situés exclusivement dans la Manche. Ce pré-traitement permet ainsi de supprimer une partie des trajectoires partiellement incomplètes, car sortant de la zone d'étude, mais reste insuffisant. En effet certaines trajectoires inféodées à des ports de la Manche sont en partie réalisées hors de la zone d'étude (Fig 2.11) et par voie de conséquence sont incomplètes et présentent des gaps (Fig 2.8). Quelle que soit l'origine des gap, afin de préserver au mieux la qualité des données VMS, il est arbitrairement décidé de conserver pour les analyses, uniquement les trajectoires contenant au plus, un gap de 3h.

Filtrage et sélection des données VMS

A partir des données VMS brutes (Table 2.4), seule une fraction des marées est conservée pour la suite de ce travail. Ici la notion de données brutes correspond aux trajectoires inféodées à la Manche (après pré-traitement).

2. S'agissant de trajectoires incomplètes nous projetons exceptionnellement les traits de côtes dont la représentation aide à la compréhension de l'origine des gaps.

TABLE 2.4 – Résumé données brutes VMS

Engin	Année	Nombre de navires	Nombre de marées	Nombre de ports
PTM	2012	44	546	19
	2013	46	636	17
OTB	2012	238	9338	102
	2013	229	9299	88

Au-delà de la sélection des marées les plus complètes (sans gaps), une attention particulière est portée sur la durée des marées. Afin de disposer d'un nombre de point suffisant à l'application des métriques descriptives des dyades, il est arbitrairement décidé de conserver uniquement les dyades présentant une durée supérieure ou égale à 10h rendant inutile la conservation des marées d'une durée inférieure à 10h. Cette étape de filtrage et de sélection des données sur leur qualité et leur durée se traduit par une suppression de plus de 20% des marées de la flottille des PTM et plus de 35% des marées de la flottille des OTB. (Table 2.5)

TABLE 2.5 – Résumé données filtrées et transformées VMS

Engin	Année	Nombre de navires	Nombre de marées	Nombre de ports
PTM	2012	38	422 (-23%)	17
	2013	45	494 (-22%)	13
OTB	2012	215	5777 (-38%)	77
	2013	217	5974 (-36%)	71

Au travers de cette étape de transformation et de sélection des données, quelques différences notables entre les deux flottilles d'intérêts peuvent être identifiées. La flottille des PTM se caractérise par un faible nombre de navires (quelques dizaines) réalisant chacun, en moyenne, une dizaine de marées chaque année. A l'inverse, la flottille des OTB apparaît beaucoup plus imposante par son effectif de navires (plus de 200 navires) réalisant en moyenne plus d'une vingtaine de marées chacun chaque année. Ainsi, c'est au travers de ces ensembles de marées que nous allons identifier et analyser les dyades.

2.3 Intensité et cohésion des relations dyadiques

Cette section peut maintenant explorer les résultats de l'application de la décomposition par les dyades des deux flottilles PTM et OTB.

2.3.1 Application aux flottilles d'études

Dans un premier temps, l'ensemble des dyades s'établissant entre les trajectoires de chaque flottille est identifié. A l'instar, de leurs effectifs et du nombre de leurs marées, les deux flottilles d'intérêts PTM et OTB se différencient grandement par le nombre de dyades que forment leurs trajectoires (Table 2.6). En effet, environ 1 500 dyades sont identifiées chaque année entre les trajectoires de la flottille PTM, tandis que l'enchevêtrement serré des trajectoires des navires OTB en forme environ 20 000. Cependant, le nombre moyen de dyades que forme chaque trajectoire, est sensiblement égal entre les deux flottilles où chaque marée, en moyenne, intervient dans la constitution de 7 à 9 dyades. On remarque également que quelle que soit la flottille, une proportion non négligeable de marées ($> 20\%$ pour les PTM, et $> 15\%$ pour les OTB) n'entre dans la constitution d'aucune dyade traduisant l'existence de marées isolées.

TABLE 2.6 – Résumé nombres de dyades

Engin	Année	Nombre de navires	Nombre de marées	Nombre de dyades	Nombre moyen dyades/marée
PTM	2012	36	330(-22%)	1118	7
	2013	43	385(-22%)	1659	9
OTB	2012	206	4823(-17%)	20734	9
	2013	205	5021(-16%)	19925	8

Si ce n'est pas au niveau du nombre de dyades qu'une éventuelle différence de comportement collectif peut être identifiée, c'est peut-être au niveau de la cohésion de ces dernières que celle-ci peut apparaître. Ainsi, dans un second temps, la cohésion de chaque dyade est évaluée. La distribution des métriques descriptives des dyades (Fig 2.12) apparaît très similaire quelle que soit la flottille avec d'un côté, une métrique de proximité dont la distribution s'apparente à une loi de puissance et de l'autre, des métriques de synchronie angulaire et de déplacement dont les distributions s'apparentent à des lois normales.

En d'autres termes, qu'il s'agisse des PTM ou des OTB, la majorité des dyades est constituée de couple de trajectoires présentant une très faible proximité simultanée (Fig 2.12a), se déplaçant en moyenne à des vitesses comparables (Fig 2.12c) et dans des directions tantôt similaires, tantôt opposées (Fig 2.12b). Ainsi décrites, les dyades sont principalement représentatives de navires se croisant ponctuellement en mer.

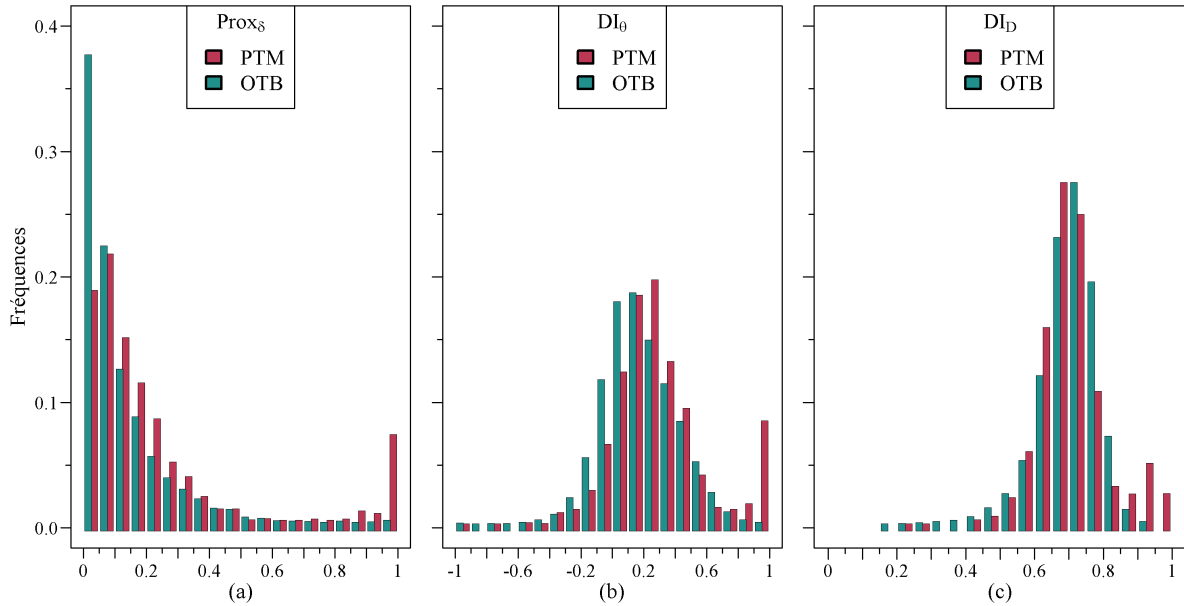


FIGURE 2.12 – Distributions des trois métriques descriptives calculée sur les dyades PTM et OTB (a) $Prox_{\delta}$, (b) DI_{θ} et (c) DI_D .

Néanmoins, contrairement aux OTB, une fraction non négligeable des dyades de la flottille PTM présente de très fortes valeurs de métriques, traduisant une très forte cohésion entre les couples de trajectoires constituant ces dyades. Ces "anomalies" sont caractéristiques du comportement de collaboration des PTM en boeufs, n'existant pas ou peu à priori au sein des OTB. C'est uniquement à ce niveau là que les distributions des deux flottilles se distinguent.

2.3.2 Classification des relations dyadiques

C'est une chose que d'identifier les dyades et de considérer indépendamment les distributions de leurs métriques, cela en est une autre que de considérer conjointement les métriques associées à chaque dyade et de les associer à différents types de comportements deux à deux. La description de l'intensité de la cohésion liant deux trajectoires par trois métriques ouvre la voie aux méthodes de classification où il s'agit ici d'être capable de discriminer les dyades à l'aide d'un modèle de mélange gaussien (GMM) [109], également décrit par Joo et al [65].

Les modèles de mélange gaussien (GMM pour "*Gaussian Mixture Model*"), supposent qu'une population peut être constituée de plusieurs sous-populations homogènes appelées

composantes. La population totale est alors un mélange de ces sous-populations présentes en proportions différentes (Fig 2.13).

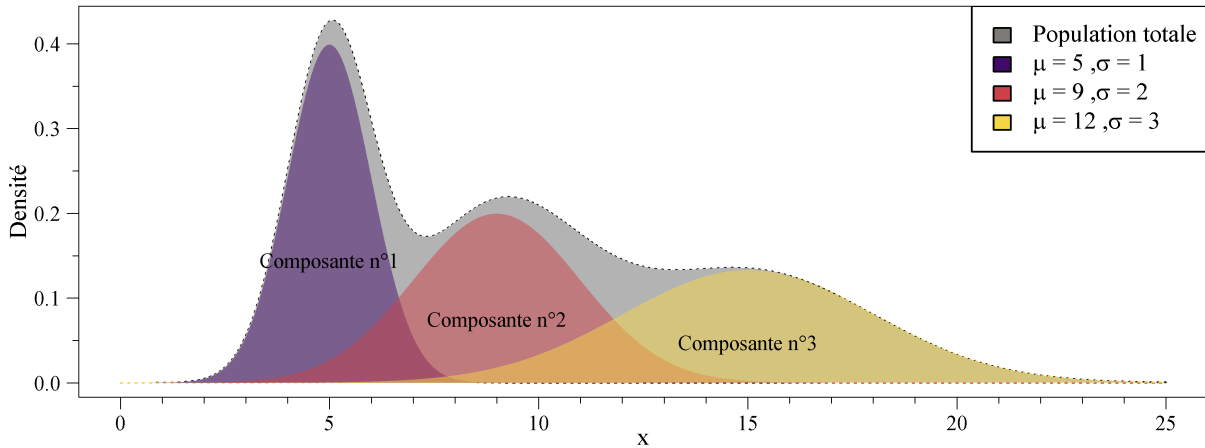


FIGURE 2.13 – Illustration d'un mélange gaussien à 3 sous-populations.

Ainsi, si $x = \{x_1, x_2, \dots, x_n\}$ est un échantillon composé de n individus appartenant à \mathbb{R}^p caractérisés par p variables aléatoires indépendantes et identiquement distribués (*iid*) issu d'un mélange de K distributions, alors sa densité f est donnée selon :

$$f(x) = \sum_{k=1}^K \Pi_k f_k(x) \quad (2.5)$$

avec, Π_k les proportions des sous-populations k et f_k les densités de la k ième distribution.

Sans entrer dans le détail, les approches par GMM sont des approches probabilistes permettant d'identifier à partir d'une population totale, les sous-populations la constituant et d'associer chaque individu à l'une ou l'autre de ces sous-populations. Il s'agit alors, dans un premier temps, d'identifier le nombre de sous-populations K , composant la population totale, et pour chacune d'entre elle, d'estimer les paramètres θ_k de sa distribution (sa moyenne μ_k et sa variance σ_k), ainsi que sa proportion (ou amplitude) π_k . Cette étape d'estimation est réalisée en vue de maximiser un critère de vraisemblance à l'aide d'une procédure itérative EM ("*Expectation-Maximisation*") [110]. Dans un second temps, il s'agit d'estimer la probabilité d'appartenance de chaque individu à chaque sous-population et d'assigner chaque individu x_i à la sous population k dont la probabilité

d'appartenance est la plus grande selon :

$$\mathbb{P}(x_i \in K_k) = \frac{\pi_k f(x_i, \theta_k)}{\sum_{j=1}^K \pi_j f(x_i, \theta_j)} \quad (2.6)$$

En termes de dyades, il s'agit de considérer que la population totale des dyades d'une flottille est constituée de sous-populations représentant chacune, une famille de dyades dont l'appartenance traduit d'un comportement 2 à 2 différent. Comme proposée par Joo et al, la classification des dyades par GMM est alors réalisée en deux temps. Dans un premier temps, la population de dyades de la flottille des PTM (toutes années confondues) est utilisée afin d'estimer les paramètres des K sous-populations avec ici, arbitrairement fixé $K = 3$. Le choix de définir trois sous-populations doit permettre de faire émerger de la classification, trois comportements deux à deux différents au regard des connaissances a priori du comportement des pêcheurs en boeufs. D'un côté, des comportements de collaboration (forte cohésion), de l'autre des comportements "indépendants" ou d'indifférences (faible cohésion) et entre les deux, des comportements ambigus traduisant l'exploitation de la même zone de pêche pouvant d'une certaine manière se rapprocher à des comportements compétitifs (moyenne cohésion). Une fois les paramètres des sous-populations ainsi estimés, dans un second temps, les dyades des deux flottilles PTM et OTB sont classées dans chacun de ces 3 groupes. Par cette approche, il s'agit de prédire à quelles sous-populations appartiennent les dyades de la flottille OTB.

Ainsi, l'ensemble des dyades est reparti en 3 groupes (sous-populations) correspondant chacun à un niveau de cohésion différent (Table 2.7). Les résultats de cette classification sont en accord avec les observations réalisées précédemment sur les différentes distributions des métriques (Fig 2.12). Pour les deux flottilles, la majorité des dyades, respectivement plus de 65% et de 75% pour les dyades PTM ou OTB, traduit une faible cohésion deux à deux (groupe n°3) représentative de rencontres ponctuelles en mer. De même, pour les deux flottilles, environ 20% des dyades traduisent une cohésion moyenne (groupe n°2).

En revanche, c'est au niveau des dyades présentant une forte cohésion (groupe n°1), caractéristiques des comportements de collaboration, que les deux flottilles se distinguent avec respectivement environ 10% et 0.5% des dyades classées dans le groupe n°1 pour les PTM et les OTB. En d'autres termes, les comportements de collaboration tels qu'ils se manifestent dans la flottille des pêcheurs en boeufs PTM sont quasiment inexistantes au sein de la flottille des chalutiers de fond OTB.

TABLE 2.7 – Classification par GMM des dyades PTM et OTB en 3 groupes de cohésion décroissante (Groupe n°1, Groupe n°2, Groupe n°3).

Engin	Année	Nombre	Total	Classification dyades		
				Groupe n°1	Groupe n°2	Groupe n°3
PTM	2012	Nbr Navires	36	32 (89%)	30 (83%)	34 (94%)
		Nbr dyades	1118	120 (11%)	213 (19%)	785 (70%)
	2013	Nbr Navires	43	40 (93%)	41 (95%)	40 (93%)
		Nbr dyades	1659	145 (9%)	416 (25%)	1098 (66%)
OTB	2012	Nbr Navires	206	16 (8%)	200 (97%)	194 (94%)
		Nbr dyades	20734	79 (0.4%)	4869 (23.5%)	15786 (76.1%)
	2013	Nbr Navires	205	21 (10%)	195 (95%)	189 (92%)
		Nbr dyades	19925	118 (0.6%)	4491 (22.5%)	15316 (76.9%)

En analysant la distribution des métriques par sous-population (Fig 2.14), concernant la flottille PTM, sans surprise, le groupe n°1 se distingue des deux autres groupes par les fortes valeurs de métrique des dyades qui le compose (Fig 2.14a-c), cette distinction apparaît également concernant les dyades du groupe n°1 de la flottille OTB (Fig 2.14d-f).

En revanche, pour les deux flottilles, la distinction entre le groupe n°2 et n°3 est bien moins marquée notamment pour les distributions des métriques de synchronie angulaire (DI_θ (Fig 2.14b,e)) et de déplacement (DI_D (Fig 2.14c,f)) qui se chevauchent largement. En effet, concernant ces deux groupes, seule la métrique de proximité ($Prox_\delta$) semble parvenir à les distinguer plus nettement.

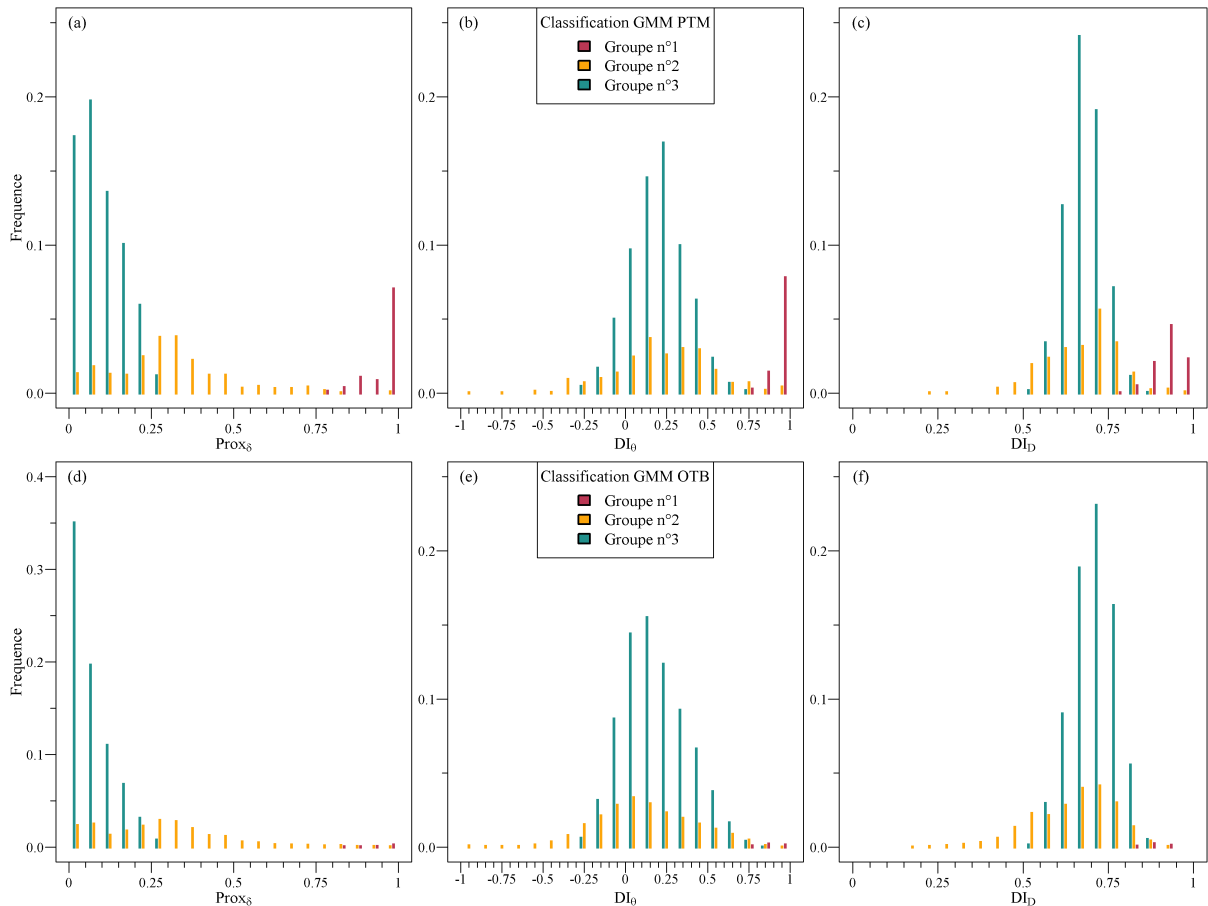


FIGURE 2.14 – Distributions relatives des métriques descriptives des dyades selon les 3 groupes de cohésion issus de la classification par GMM pour les deux flottilles PTM et OTB

Enfin, les résultats de cette classification peuvent être visualisés au travers d'une analyse par composante principale (ACP) [111] permettant cette fois-ci une représentation prenant en compte conjointement les trois métriques descriptives (Fig 2.15). Dans ces ACP, les individus sont les dyades et les variables sont les trois métriques descriptives qui leurs sont associées. Là encore, les projections de l'ACP soutiennent les interprétation précédentes où, concernant les dyades de la flottille PTM, la projection présente un groupe n°1 bien discriminé tandis que les groupes n°2 et n°3 se confondent (Fig 2.15a) Concernant la projection des dyades de la flottille OTB, la discrimination du groupe n°1 est moins nette tandis que les groupes n°2 et n°3 forment un continuum (Fig 2.15b).

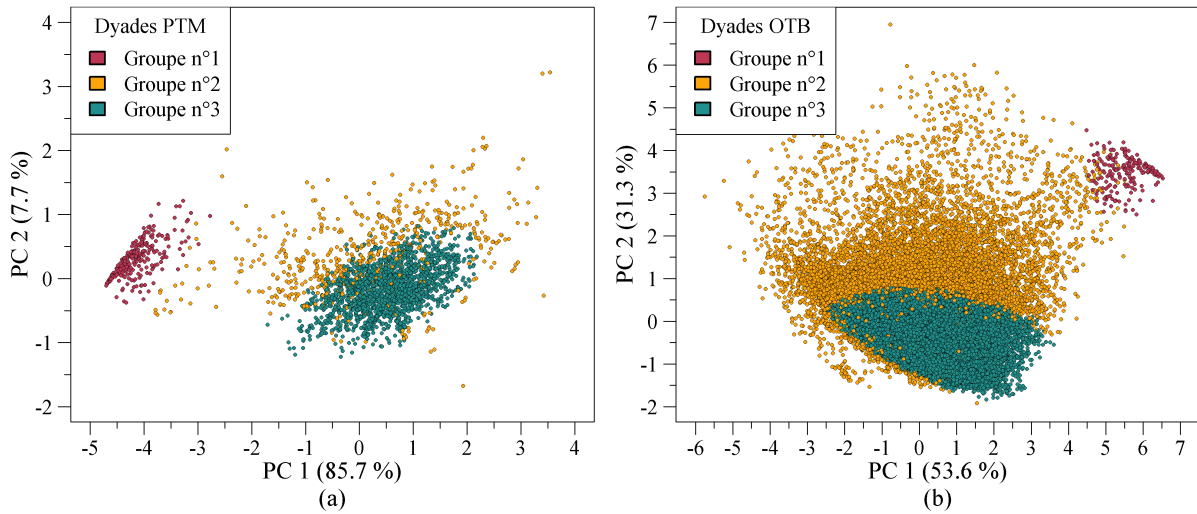


FIGURE 2.15 – Analyse par composantes principales des dyades (a) PTM et (b) OTB selon leurs trois métriques descriptives et selon la classification par GMM.

2.4 Des relations 2 à 2 aux relations n à n

A ce stade, sur la base de travaux antérieurs, il a été possible de décomposer un ensemble de trajectoires en un ensemble de relations dyadiques dont l'intensité de leur cohésion est interprétée comme différents types de comportements deux à deux. Afin d'interroger le comportement collectif régnant au sein d'une flottille, il est nécessaire de rétablir les liens entre l'ensemble des relations dyadiques afin de pouvoir interroger leur organisation.

Au cours du chapitre précédent, nous avons vu qu'il existe différents types de graphes permettant de traduire différents phénomènes. Dans cette section, nous montrons au travers de l'intégration des relations dyadiques, qu'un même phénomène peut être traduit en différents types de graphes.

2.4.1 Le graphe de flottille, un objet modulable

Dans le cadre de ce travail, une flottille est un ensemble de marées dont l'existence d'une dyade traduit l'existence d'une relation entre deux d'entre elles. Les dyades peuvent alors être utilisées pour construire un graphe afin de visualiser à l'échelle d'une flottille l'organisation d'un ensemble de relations deux à deux, en d'autres termes, l'émergence des relations n à n .

Considérons par exemple l'ensemble des trajectoires de la flottille PTM enregistrées sur deux semaines d'observation (Fig 2.6a). Après identification et classification des dyades réalisées au sein de cet ensemble de trajectoires, quatre graphes peuvent être construits (Fig 2.16), un pour chaque groupe de comportement (Fig 2.16a-c) et un contenant l'ensemble des dyades (Fig 2.16d).

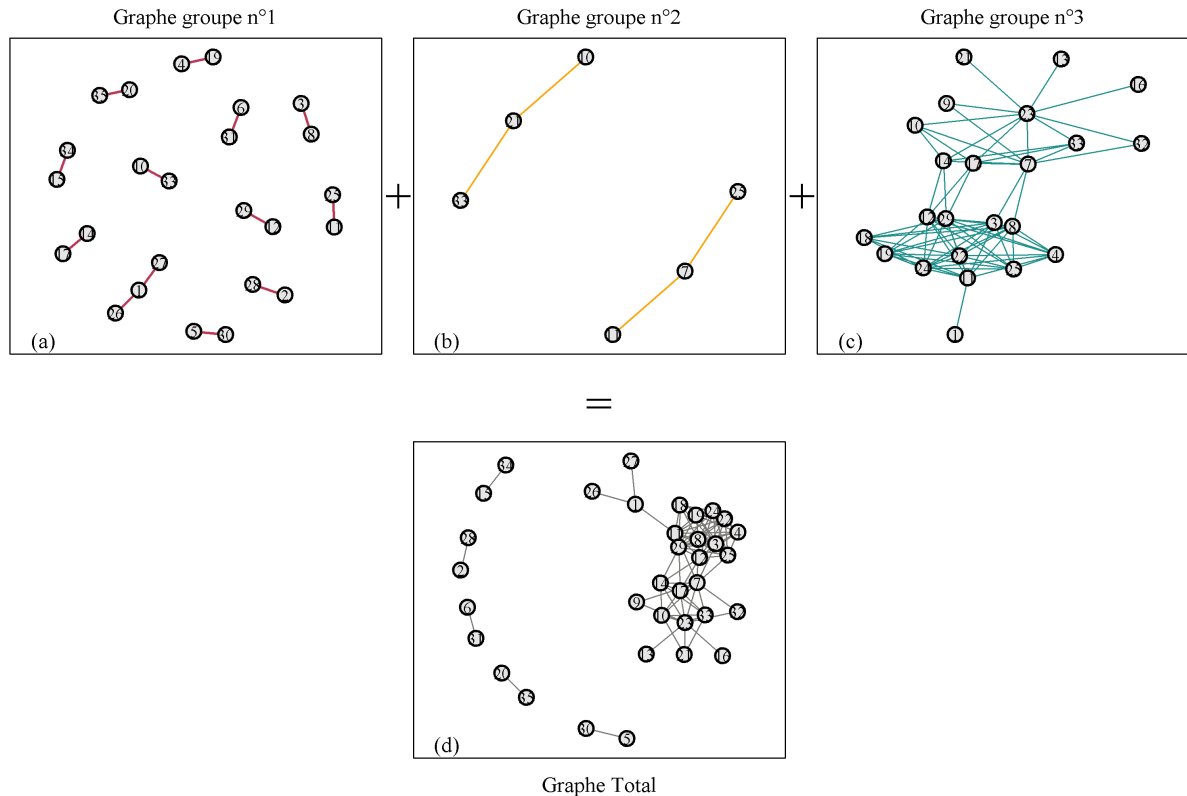


FIGURE 2.16 – Illustration de l'intégration sous la forme de graphes d'un ensemble de dyades PTM (a) du groupe n°1, (b) du groupe n°2, (c) du groupe n°3 et (d) totale.

Il s'agit ici de graphes simples non orientés où chaque noeud correspond à une marée et l'existence d'une arête entre deux noeuds traduit l'existence d'une dyade entre leur trajectoires respectives. Ainsi, dans cet exemple, les dyades du groupe n°1 conduisent à la formation d'un graphe où les noeuds sont organisés en couples isolés les uns des autres (Fig 2.16a) (et d'un triplet), traduisant l'exclusivité de la relation de collaboration. A l'inverse, les dyades du groupe n°3 conduisent à la formation d'un graphe où les arêtes forment un enchevêtrement (Fig 2.16c) à l'image des trajectoires, traduisant des multiples croisements en mer que réalisent l'ensemble des navires.

Finalement, les trois graphes relatifs aux trois groupes de comportement peuvent être

additionnés (combinés) sous la forme d'un unique graphe (Fig 2.16d) contenant l'intégralité des dyades, rendant compte de l'ensemble des relations dyadiques s'établissant entre toutes les marées. Notons que les marées isolées, c'est-à-dire celles n'ayant contribué à aucune dyade ne sont pas représentées ici, dans le cas contraire elles apparaîtraient sous la forme de noeuds isolés.

Plutôt que de représenter les liens entre les marées, il est également possible de représenter les liens entre les navires. En effet, sur deux semaines d'observation, il est possible qu'un même navire réalise plusieurs marées au travers desquelles il peut former plusieurs dyades. Représenter les liens entre les navires consiste simplement à représenter l'ensemble des dyades qu'il a formé au cours de ces multiples marées (Fig 2.17b)

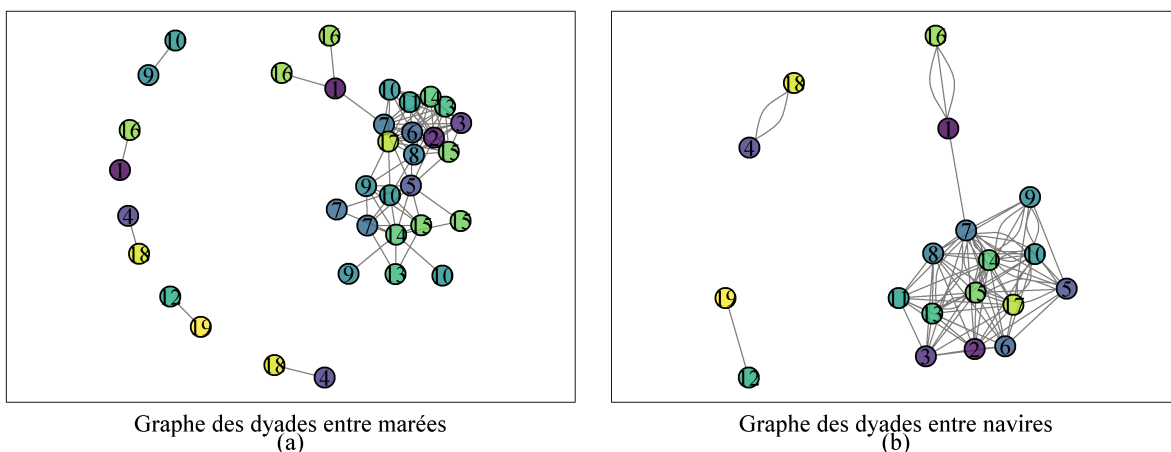


FIGURE 2.17 – Représentation des graphes illustrant les dyades (a) entre les marées (trajectoires) ou (b) entre les navires.

Dans cette seconde représentation, il s'agit d'un graphe multiple où l'existence de plusieurs arêtes entre un même couple de noeuds traduit l'existence de plusieurs dyades. Les arêtes multiples, en mettant en évidence la récurrence des relations dyadiques entre un même couple de navires, peuvent être la traduction d'une forme de fidélité. Sur cet exemple, le navire n°16 a réalisé trois marées au cours desquelles, il a formé exclusivement des dyades avec le navire n°1 ayant quant à lui réalisé deux marées.

2.4.2 Le graphe de flottille, un objet temporel

Derrière leur aspect statique, les graphes des relations dyadiques possèdent une composante temporelle relative à la date et à l'heure de la formation des dyades (Fig 2.18).

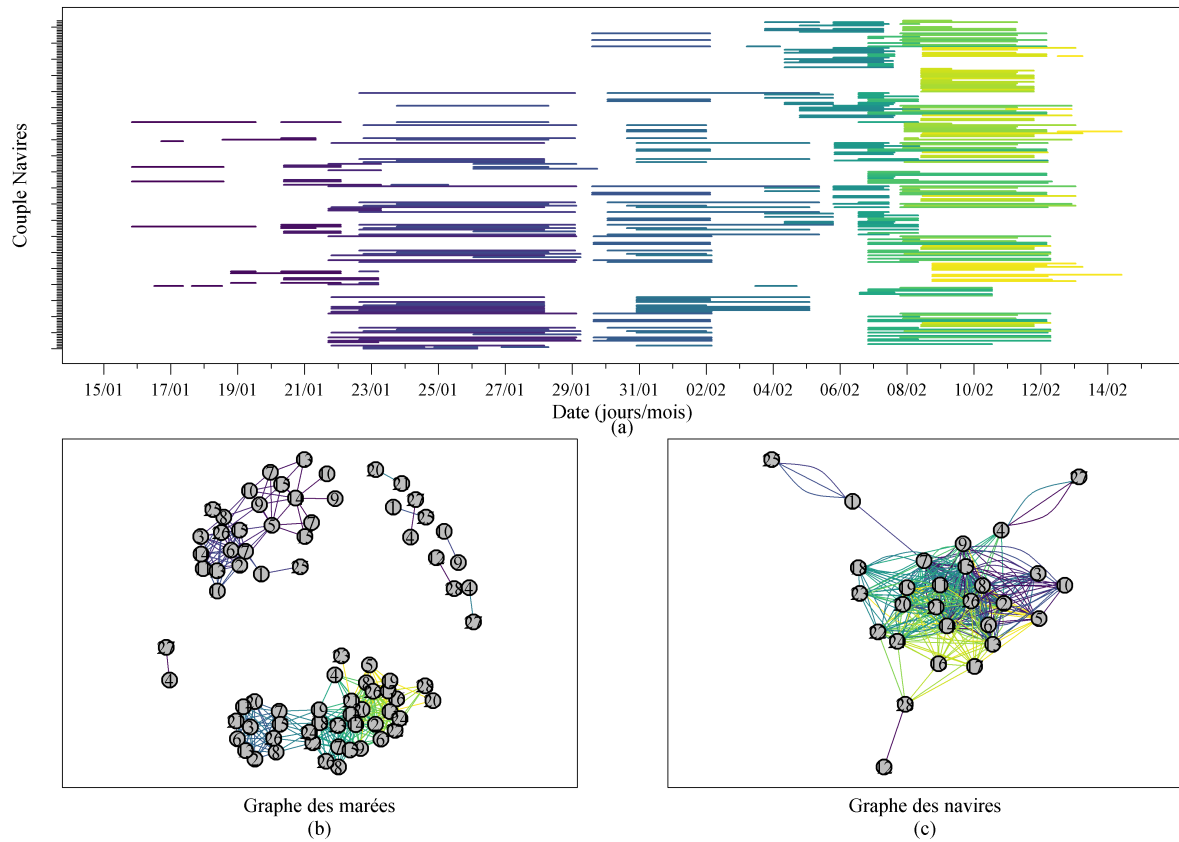


FIGURE 2.18 – Représentation de la dynamique temporelle des dyades au sein des graphes (a) des marées et (b) des navires.

Un code couleur permet en effet de rendre compte que l'organisation des dyades au sein d'un graphe est fonction du temps et, notamment, lorsqu'il s'agit de représenter des graphes de marées. En effet, les marées n'ont la possibilité de former une dyade qu'avec les autres marées s'établissant dans le même espace temporel, rendant impossible l'établissement d'une dyade avec une marée déjà terminée (du passé) ou qui n'a pas encore commencé (du futur). A l'échelle des graphes mettant en relation les marées (Fig 2.18a), cette contrainte temporelle se traduit par la formation de groupes de noeuds (de marées) traduisant dans ce cas précis une discontinuité temporelle. Cette composante temporelle est en revanche indécélable dans l'organisation des graphes mettant en relation les navires (Fig 2.18b) à condition que ces derniers réalisent régulièrement des marées.

2.4.3 Le graphe de flottille, un objet spatial

De la même manière que la dynamique temporelle des dyades est implicitement retranscrite dans les graphes, ces derniers sont également sous fonction d’une dynamique spatiale. En effet, du fait que chaque marée soit définie par un port de départ et un port de retour, conditionnant son rayon d’action, et que la constitution d’une dyade repose sur un critère de proximité physique, l’organisation des dyades les unes par rapport aux autres reflète la répartition spatiale des trajectoires (Fig 2.19b).

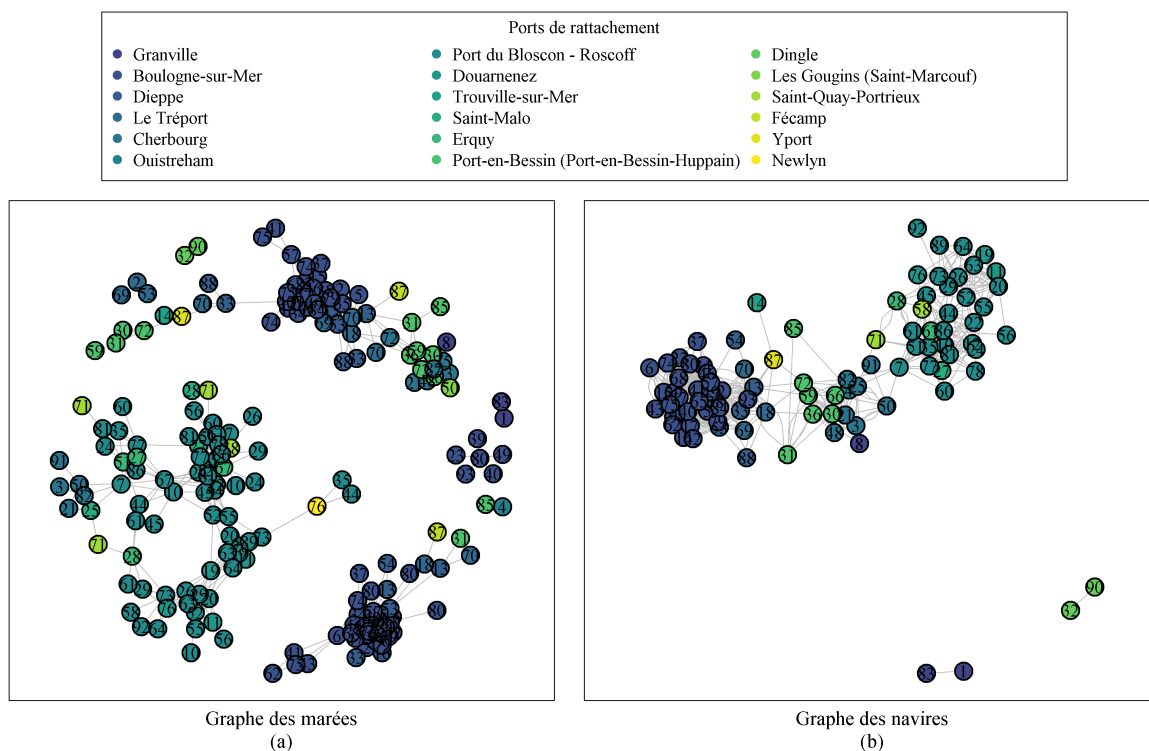


FIGURE 2.19 – Représentation de la composante spatiale des dyades au sein des graphes (a) des marées et (b) des navires.

Ce point peut être illustré au travers des graphes obtenus sur l’ensemble de trajectoires de la flottille OTB enregistré sur deux semaines d’observations (Fig 2.6b) (les navires OTB présentant une plus grande répartition spatiale que les navires PTM), en représentant par un code couleur, pour chaque noeud le port de départ (Fig 2.19). Cette fois-ci, qu’il s’agisse des graphes des marées ou des graphes des navires, la composante spatiale est bien visible dans l’organisation du graphe, par la formation de groupe de noeuds présentant le même port d’origine. Évidemment, concernant le graphe des marées (Fig 2.19a) une composante

temporelle est également présente se reflétant par la formation de plusieurs groupes de noeuds issus d'un même port. En considérant un temps d'observation suffisamment long et sous l'hypothèse que chaque navire reste rattaché à un même port, il peut être supposé que l'organisation du graphe des navires traduirait un gradient Est-Ouest relatif à la répartition des ports le long de la Manche.

Au travers de ces quelques exemples de graphes, il est question d'illustrer d'une part les multiples façons dont peuvent être intégrées des relations 2 à 2, et d'autre part, la richesse en information que peuvent contenir les graphes malgré leur apparente "simplicité". Ainsi, les graphes par leur modularité laissent le choix de la représentation à adopter, dépendante de l'information à mettre en avant.

Pour la suite, s'agissant d'interroger l'indépendance des trajectoires, seuls les graphes de marées seront considérés.

2.5 Conclusion du chapitre

Dans ce chapitre, nous avons décrit une approche permettant de décomposer un ensemble de trajectoires reconstruit sur la base des données VMS en un ensemble de relations dyadiques. Nous avons montré que l'analyse des dyades peut permettre d'interpréter certains comportements deux à deux et notamment les comportements de collaboration caractéristiques de la flottille des PTM. Concernant la flottille des OTB, cette même analyse a montré que ces comportements de collaboration de "type PTM" étaient très faiblement représentés, ne permettant pas de présupposer de l'existence d'un quelconque comportement collectif.

Au travers de l'intégration des relations dyadiques sous la forme d'un graphe nous avons montré différentes manières de révéler l'organisation des dyades à l'échelle d'une flottille. Nous pouvons désormais nous pencher sur l'analyse de cette organisation avec, pour le chapitre suivant, la confrontation des graphes de flottilles à des graphes d'indépendance du modèle d'Erdős-Rényi.

DE L'INDÉPENDANCE DES TRAJECTOIRES À L'INDÉPENDANCE DES RELATIONS DYADIQUES

Au cours du chapitre précédent après avoir dénoué un ensemble de trajectoires et l'avoir fragmenté en un ensemble de dyades, nous les avons ré-associés sous la forme d'un graphe restaurant ainsi, au travers des dyades, les liens unissant chacune des trajectoires.

Dans ce chapitre nous proposons le développement d'une approche permettant d'interroger les propriétés d'organisation des arêtes au sein d'un graphe. Articulée autour du questionnement de l'indépendance des relations dyadiques des graphes de flottilles, cette approche dépasse les frontières de l'halieutique en s'inscrivant dans une problématique intrinsèquement liée à la théorie des graphes, la comparaison de graphes.

3.1 Système d'hypothèse et démarche :

Dans ce manuscrit, il est question d'interroger la manière dont des navires de pêche se déplacent mutuellement en mer, avec pour question centrale, sont-ils indépendants les uns des autres. Sous cette hypothèse d'indépendance, il est considéré que le comportement de chaque navire ne dépend pas du comportement des autres navires. Si cette hypothèse d'indépendance s'avère exacte, alors elle devrait se traduire dans la manière dont les navires se rencontrent en mer, c'est-à-dire dans la manière dont les navires forment des dyades.

L'intégration des relations dyadiques sous la forme d'un graphe permet de visualiser l'organisation des rencontres en mer à l'échelle d'une flottille. Dans ce chapitre nous proposons d'interroger les propriétés d'organisation des dyades par l'intermédiaire de méthodes de comparaison de graphes. Cette approche repose sur l'utilisation de modèles de

graphe dont les réalisations répondent à des propriétés d'organisation connues. En d'autres termes, l'identification de similarités entre des graphes de flottille et des réalisations de modèles de graphe permettrait d'en déduire les propriétés d'organisation des relations dyadiques. Dans cette approche, le modèle d'Erdős-Rényi, caractérisée par la propriété d'indépendance des arêtes, occupe une place centrale. C'est sur la base de cette réflexion qu'un nouveau système d'hypothèse est construit établissant le lien entre l'indépendance des trajectoires et l'indépendance des relations dyadiques :

Systeme d'hypothèse n°2

Hypothèse : Si les trajectoires sont indépendantes, alors les rencontres en mer (les dyades) devraient l'être également.

Prédiction : Les trajectoires de la flottilles des PTM ne sont pas indépendantes, leur rencontre en mer, ne le sont alors également pas.

Expérimentation : Comparaison des graphes de la flottille PTM à des graphes d'indépendance Erdős-Rényi.

Observation : Voir résultats section 3.3.3

3.2 La similarité entre graphes

Nous l'avons vu, derrière les modèles de graphes, se trouve en toile de fond, l'idée de faire ressembler les réalisations de ces modèles aux graphes du monde réel, mais qu'est-ce que cela signifie au bout du compte que deux graphes se ressemblent ou qu'ils ne se ressemblent pas ? Dans cette section, nous nous intéresserons à la question de la similarité entre graphes en interrogeant d'une part ce qu'elle signifie, et d'autre part comment au travers de différentes méthodes, cette similarité peut être évaluée.

3.2.1 Quelle similarité ?

Le sujet de la comparaison de graphes qu'il s'agisse de comparer entre eux des réalisations de modèles de graphes ou des graphes empiriques, est motivé par le même objectif, "*identifier et quantifier les différences et/ou les similarités qui existent entre un ensemble de graphes*" [112]. Toutefois, malgré cet objectif commun derrière toutes les méthodes de comparaison de graphes, il n'existe pas une manière mais des manières de comparer des graphes.

En effet, l'utilisation des graphes s'étant insinuée peu à peu dans tous les champs de la recherche (en mathématiques, en physique, en biologie, en médecine, en écologie, en sociologie, etc...), le problème de la comparaison de graphe est apparue comme étant une réponse à des besoins différents, où il s'agit de mettre en avant certaines formes de similarité ou de différence. En d'autres termes, ce n'est pas la même similarité qui est testée lorsqu'il s'agit de comparer par exemple des graphes de protéines, que celle lorsqu'il s'agit de comparer des graphes de réseaux sociaux.

Pour répondre à ces besoins différents, une multitude de méthodes de comparaison a vu le jour dans le but d'exacerber un critère de similarité ou de différence propre à chaque discipline. Certains travaux ont essayé de classer ces différentes méthodes de comparaison en de grands ensembles, par exemple en fonction de l'échelle d'analyse sur laquelle repose la comparaison, soit, les comparaisons à micro-échelle (portant sur les caractéristiques des noeuds), à méso-échelle (portant sur l'organisation de groupes de noeuds), ou encore à macro-échelle (portant sur les caractéristiques de l'ensemble du graphes) [113]. Toutefois, en l'absence de classification standardisée et faisant consensus, il est choisi ici, de passer en revue pêle-mêle quelques-unes des méthodes de comparaison de graphes les plus courantes et qui interroge la similarité des graphes sous des aspects différents.

Les méthodes de comparaison sont souvent matérialisées sous la forme d'une "distance" entre graphes, où une faible distance indique une forte similarité et inversement. Néanmoins, cette appellation de distance est souvent trompeuse car peu de méthodes de comparaison de graphes présentent les propriétés d'une distance au sens mathématique et notamment, le respect de l'inégalité triangulaire.

Les différentes méthodes de comparaisons présentées ci-dessous seront illustrées, si elles le permettent, au travers d'une comparaison simple devant illustrer d'une part, le principe de fonctionnement de la comparaison, et d'autre part, le résultat de cette comparaison. Ainsi, trois graphes seront utilisés en guise d'illustration, une version simple et non-orientée du graphe de Sampson (Fig 3.1a) notée G_a , une version altérée de ce dernier (Fig 3.1b) notée G_b issue de la suppression de deux arêtes liant la communauté des "Jeunes Turcs" à celle de "l'Opposition Loyale", et enfin un graphe G_c d'Erdős-Rényi contenant le même nombre de noeuds N et d'arêtes m que la version simple et non orientée du graphe de Sampson (Fig 3.1c) soit $G_c = G_{ER}(N = 18, m = 41)$.

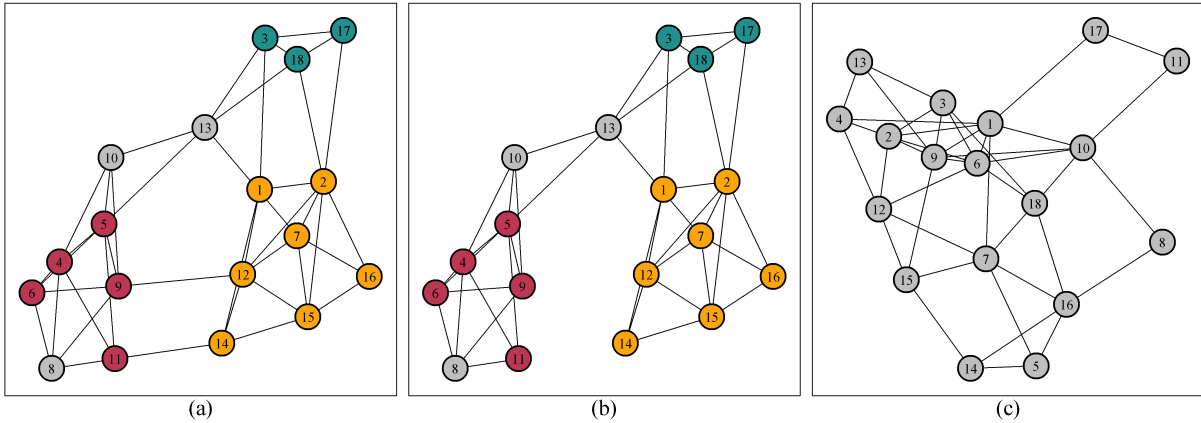


FIGURE 3.1 – Les trois graphes d’illustration des méthodes de comparaison de graphes (a) G_a le graphe de Sampson, (b) G_b le graphe de Sampson modifié et (c) G_c une réalisation du modèle de Erdős-Rényi.

Deux comparaisons seront effectuées, d’un côté la comparaison entre les deux versions du graphe de Sampson (G_a et G_b) et de l’autre une comparaison entre le graphe de Sampson G_a et le graphe d’Erdős-Rényi G_c . Il est attendu des différentes méthodes de comparaison qu’elles attribuent une plus faible distance entre les deux versions du graphe de Sampson.

3.2.2 L’isomorphisme, une similarité binaire

La première question qui se pose lors de la comparaison entre deux graphes est de savoir si ces deux graphes sont identiques ou non, ou isomorphes[114] en termes de théorie des graphes. Ainsi, deux graphes G_1 et G_2 sont dits isomorphes, noté $G_1 \simeq G_2$, s’il existe une application bijective $f : V(G_1) \rightarrow V(G_2)$ entre leurs deux ensembles de noeuds. Selon cette bijection, pour tout couple de noeuds $\{v_i; v_j\}$ adjacents dans G_1 , alors $f(v_i)$ et $f(v_j)$ sont adjacents dans G_2 .

Tester l’isomorphisme de deux graphes reste toutefois une tâche relativement limitée. D’une part le résultat est binaire, soit les deux graphes sont isomorphes, soit ils ne le sont pas sans pour autant pouvoir spécifier en quoi ils diffèrent. D’autre part, l’isomorphisme ne peut être recherché qu’entre deux graphes de même ordre $N_1 = N_2$ et de même taille $m_1 = m_2$, interdisant toute comparaison entre graphes d’ordre et/ou de taille différentes (les graphes G_a et G_b ne sont naturellement pas isomorphe). Enfin, tester l’isomorphisme de deux graphes fait intervenir des algorithmes souvent lourds dont le temps de réponse

est incertain [115] et n'est pas connu pour être solvable dans un temps polynomiale, symptôme de la "*maladie de l'isomorphisme*" [116].

Ainsi la recherche de l'isomorphisme reste une approche restreinte à certaines applications bien spécifiques, et souvent bien éloignées des problématiques rencontrées en écologie, en biologie, en sociologie ou dans le travail développé ici.

3.2.3 Les "Graph Edit Distances" (GED)

Les "*Graph Edit Distances*" (GED) qui pourrait être traduit par "distances d'édition entre graphes", où la notion d'édition est synonyme de modification/transformation, représentent l'une des plus anciennes méthodes de comparaison de graphes parue pour la première fois au début des années 1980 [117]. Cette distance repose sur un principe extrêmement simple, considérant deux graphes G_1 et G_2 , quelles sont les modifications ou transformations à effectuer sur G_1 pour le transformer en G_2 . Autrement dit, si G_1 et G_2 sont différents, quelle est la manière la plus optimale de rendre ces deux graphes isomorphes.

Selon cette approche, il s'agit de recenser l'ensemble des transformations, appelées "opérations d'édition" comme l'ajout, la suppression ou la permutation d'une arête ou d'un noeud, nécessaires à appliquer sur G_1 afin de le transformer en G_2 . La transformation de G_1 vers G_2 est alors traduite sous la forme d'une séquence d'opérations d'édition, notée " s ". Par exemple, la transformation optimale du graphe G_a vers G_b est une séquence contenant deux opérations d'édition relatives à l'ajout de deux arêtes entre les couples de noeuds $\{\{9, 12\}$ et $\{11, 14\}\}$, ou bien la séquence inverse, la suppression de ces deux arêtes dans le cas de la transformation G_b vers G_a .

Comme il n'existe pas une solution unique pour transformer G_1 en G_2 , la complexité de cette approche réside dans l'identification de la séquence d'opérations d'édition **optimale** parmi toutes les séquences possibles s_1, s_2, \dots, s_n . Pour ce faire, chaque type d'opération d'édition (ajout, suppression, permutation) est associé à un **coût** permettant de quantifier le coût total de chaque séquence d'édition comme la somme des coûts individuels de chaque opération. La séquence optimale notée s' est celle minimisant le coût total selon :

$$s' = \operatorname{argmin}_{s_j} \{c(s_i) | 1 \leq i \leq n\} \quad (3.1)$$

avec $c(s_i)$ la fonction du coût de chacune des opérations effectuées. La distance (GED)

entre deux graphes est alors le coût total de la séquence optimale :

$$d_{GED}(G_1, G_2) = c(s'). \quad (3.2)$$

Avec par exemple, pour $d_{GED}(G_a, G_b) = 2$ si le coût de l'ajout d'une arête est fixé à 1.

A partir de cette idée simple, cette distance GED initialement introduite par Sanfiulu et Fu [117] en 1983 a connue une véritable évolution, notamment en la rendant applicable à la fois pour des graphes orientés ou non-orientés, mais également par l'évolution de la fonction des coûts à attribuer à chaque opération d'édition, ou enfin par l'amélioration des algorithmes générant les séquences d'éditions [118, 119].

Ainsi cette distance apparaît particulièrement intéressante lorsqu'il s'agit de suivre l'évolution d'un phénomène au travers d'une série chronologique de graphes contenant les mêmes ensembles de noeuds, en permettant le suivi de la transformation d'un graphe au cours du temps. Elle apparaît également particulièrement sollicitée dans les problématiques d'identification des structures communes entre deux graphes [120] ou bien dans les algorithmes de recherche d'image [121]. Néanmoins cette distance ne permet pas véritablement d'interroger la topologie des graphes et encore moins leurs propriétés. De plus, malgré des tentatives visant à normaliser cette distance, elle reste toujours sensible aux différences d'ordre N et de taille m entre les graphes à comparer [122]. Cette distance ne sera pas utilisée dans ce manuscrit.

3.2.4 Les "features-based distances"

Les "*features-based distances*" qui pourrait être traduit par "distance des caractéristiques entre graphes" est un type d'approche répondant à la problématique de la comparaison de graphes. Cette approche repose sur un principe tout aussi simple que la précédente, tout graphe peut être décrit au travers d'un ensemble de caractéristiques ("*features*") rendant compte de son organisation topologique. Certaines de ces caractéristiques ont déjà été évoquées précédemment, comme par exemple, la distribution des degrés, le degré moyen, le coefficient de clustering ou le diamètre, mais il en existe encore bien d'autres, comme la centralité, la modularité, la distribution des cliques, des motifs, des graphlets, etc. Ainsi, pour ce type de distance, deux graphes G_1 et G_2 peuvent être comparés par l'intermédiaire de la comparaison de leurs caractéristiques topologiques.

Parmi les features-based distance, la distance NetSimile [123] est peut être l'une des

plus abouties. Pour des raisons de simplicité il ne sera pas question ici de présenter cette distance telle qu'elle a été initialement proposée, mais d'en présenter une version simplifiée permettant tout de même d'en saisir son principe. Dans cette version simplifiée, la construction de la distance est identique à celle proposée à l'origine, la différence se situe au niveau des caractéristiques utilisées pour calculer cette distance.

A titre d'illustration, trois caractéristiques sont considérées ici, le degré et la transi-
 tivité qui ont déjà été définis précédemment et la centralité ("*closness centrality*") qui se
 définit selon :

$$C_B(v_i) = \frac{1}{\sum_{j \neq i} d(v_i, v_j)} \quad (3.3)$$

comme étant, pour un noeud v_i l'inverse de la somme de tous les plus courts chemins
 entre le noeud v_i et chaque noeud v_j . Cette métrique renseigne comme son nom l'indique,
 pour chaque noeud à quel point ce dernier occupe une place centrale au sein d'un graphe.

Ainsi pour un graphe G d'ordre N , à partir de ces $k = 3$ caractéristiques, une matrice
 $N \times k$ est construite, contenant N lignes (une pour chaque noeud) et k colonnes (une pour
 chaque caractéristique) ou chaque ligne i contient pour le noeud v_i ses k caractéristiques.
 La matrice des caractéristiques d'un graphe G est ensuite résumée en l'agrégeant sous
 la forme d'un vecteur signature \vec{s}_G contenant un résumé de l'information des k carac-
 téristiques. Pour ce faire, pour chaque colonne k considérée comme autant de variables
 aléatoires X_k , sont définis sa médiane (med_k) ainsi que ses moments d'ordre un à quatre
 soit sa moyenne (μ_k), son écart-type (σ_k), son asymétrie ("*skewness*") (γ_k) et son coefficient
 d'acuité ("*kurtosis*") (β_k) définis selon :

$$\text{la moyenne} \quad \mu_X = \frac{1}{N} \sum_{i=1}^N X_i \approx \mathbb{E}(X) \quad (3.4)$$

$$\text{l'écart-type} \quad \sigma_X = \sqrt{\mathbb{E}[(X - \mu(X))^2]} \quad (3.5)$$

$$\text{l'asymétrie} \quad \gamma_X = \mathbb{E} \left[\left(\frac{X - \mu_X}{\sigma_X} \right)^3 \right] \quad (3.6)$$

$$\text{l'acuité} \quad \beta_X = \mathbb{E} \left[\left(\frac{X - \mu_X}{\sigma_X} \right)^4 \right] \quad (3.7)$$

Et afin de remplir le vecteur signature $\vec{s}_G = [med_1, \mu_1, \sigma_1, \gamma_1, \beta_1, \dots, med_3, \mu_3, \sigma_3, \gamma_3, \beta_3]$
 (Fig 3.2).

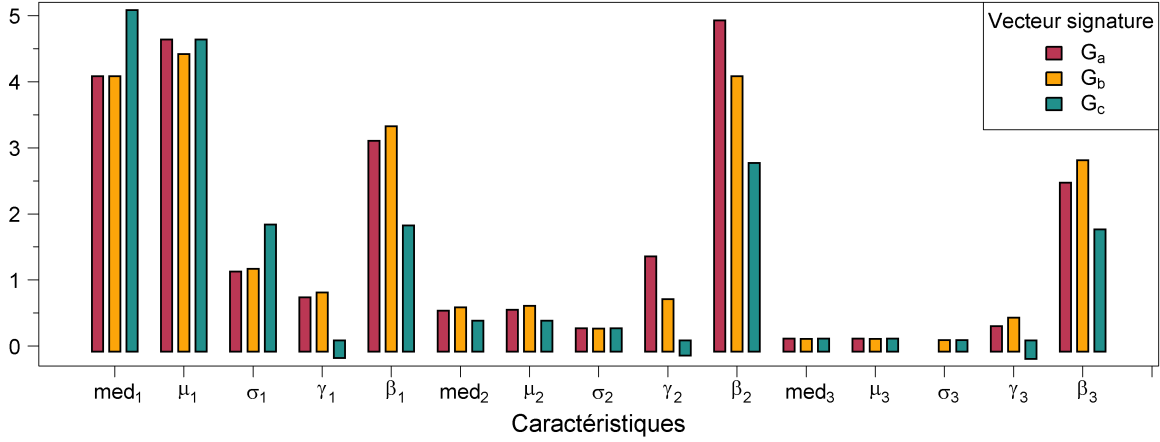


FIGURE 3.2 – Représentation des vecteurs signatures des trois graphes d’illustration G_a , G_b et G_c définies sur la base de trois métriques topologiques, le degré (1), la transitivité (2) et la centralité (3).

La distance NetSimile entre deux graphes G_1 et G_2 est alors définie comme la distance de Canberra[124] entre leur deux vecteurs signatures \vec{s}_{G_1} et \vec{s}_{G_2} selon :

$$d_{NetSimile}(G_1, G_2) = \sum_{i=1}^d \frac{|\vec{s}_{G_1}[i] - \vec{s}_{G_2}[i]|}{|\vec{s}_{G_1}[i]| + |\vec{s}_{G_2}[i]|}, \quad (3.8)$$

et repose sur la comparaison fine des distributions d’un ensemble de caractéristiques topologiques. Cette distance présente l’avantage d’être extrêmement modulable selon le choix des métriques résumant le graphe. Néanmoins cette distance peut souffrir de la dépendance de certaine métrique topologique à l’ordre N et/ou à la taille m du graphe.

Considérons par exemple les réalisations du modèle d’Erdős-Rényi $G_{ER}(N, m)$ avec $N = 1000$ et m croissant. Dans le modèle d’Erdős-Rényi $G_{ER}(N, p)$ la croissance de la taille m d’un graphe peut être suivie au travers du paramètre p , tandis que dans le modèle $G_{ER}(N, m)$ le suivi de la croissance se fait au travers de la densité du graphe, soit le rapport entre le nombre d’arêtes m et le nombre de couples de noeuds possible, selon $d = \frac{2m}{N(N-1)}$, avec $d = 0$ le graphe est vide et $d = 1$ le graphe est complet. Ainsi, il est possible d’observer la dépendance à la densité de certaines métriques topologiques comme par exemple celle du degré (Fig 3.3).

Le degré médian et moyen (Fig 3.3a,b) sont par exemple corrélés linéairement à la densité du graphe, de même l’évolution de l’écart-type (Fig 3.3c) présente une forme parabolique, symétrique au regard de la densité et centrée sur la densité médiane $d = 0.5$.

Ainsi, dans le cadre d'une comparaison entre deux graphes de densités différentes, ces caractéristiques rendront compte de la différence de densité sans pour autant apprécier la similarité des propriétés topologiques des graphes, ici l'organisation indépendante de leurs arêtes. Cette influence de la densité se retrouve également dans la mesure de l'asymétrie de la distribution des degrés (Fig 3.3d), en tendance positive à faible densité ($d < 0.5$) et négative à forte densité ($d > 0.5$) rendant compte du "déplacement" de la distribution des degrés de la "gauche" vers la "droite" s'accompagnant d'une inversion de côté de la queue de distribution. Il semble finalement que seule l'acuité de la distribution des degrés soit insensible à la variation de la densité (Fig 3.3e).

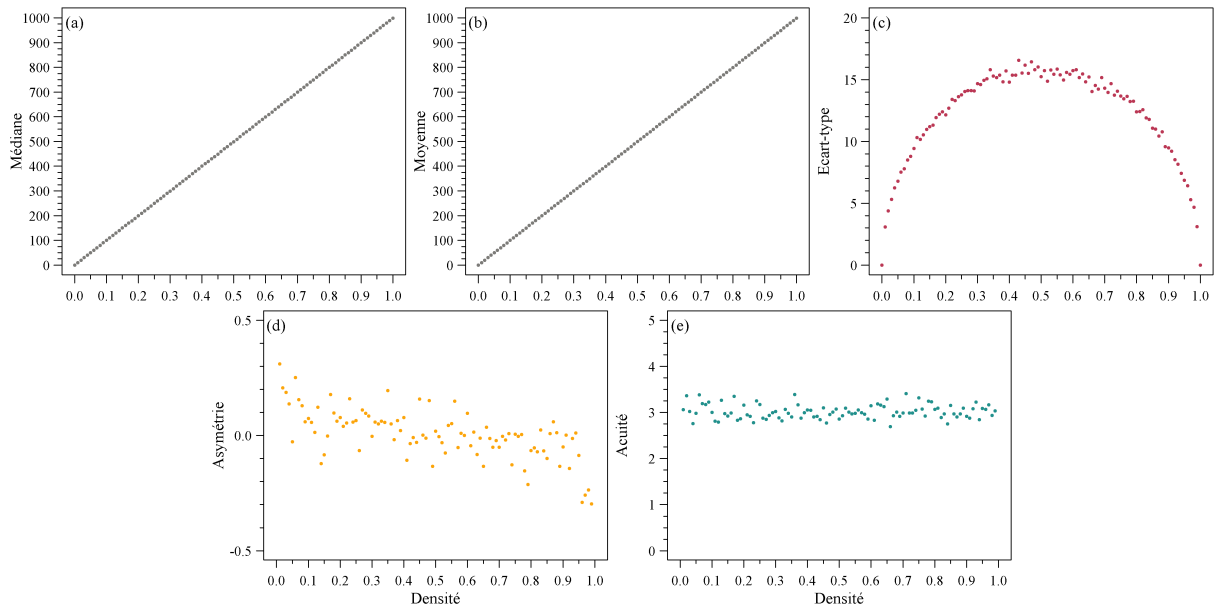


FIGURE 3.3 – Évolution des caractéristiques de la distribution des degrés d'un graphe d'Erdős-Rényi $G_{ER}(N, m)$ de densité croissante, (a) sa médiane, (b) sa moyenne, (c) son écart-type, (d) son asymétrie et (e) son acuité.

Cette influence de la densité mais également de l'ordre sur les métriques topologiques peut rendre critique l'utilisation des distances basées sur ces métriques, notamment lorsqu'il s'agit de rechercher des similarités de propriétés topologiques (indépendance des arêtes, organisation scale-free, etc.) entre graphes.

3.2.5 Les distances spectrales

Les distances spectrales découlent naturellement du fait que les graphes peuvent être représentés sous la forme d'une matrice d'adjacence, ouvrant ainsi la voie aux différentes

analyses matricielles. Le terme de distance spectrale fait ainsi référence à un ensemble de distances s'appuyant sur la décomposition des matrices d'adjacences ou des matrices qui en dérivent, en **valeurs propres** et vecteur propres associés [125].

Soit un graphe G et A sa matrice d'adjacence où on rappelle que dans le cas d'un graphe simple non orienté cette dernière est une matrice symétrique et binaire :

$$a(i, j) = \begin{cases} 1 & \text{si } v_i \text{ et } v_j \text{ sont adjacents} \\ 0 & \text{sinon} \end{cases} \quad (3.9)$$

Comme toute matrice carrée, cette matrice A peut être décomposée en une suite de valeurs propres λ_k^A positives, négatives ou nulles et contenues sous la forme d'un vecteur de leurs amplitudes décroissantes $\{\lambda_1^A \geq \lambda_2^A \geq \dots \geq \lambda_N^A\}$ (Fig 3.4a).

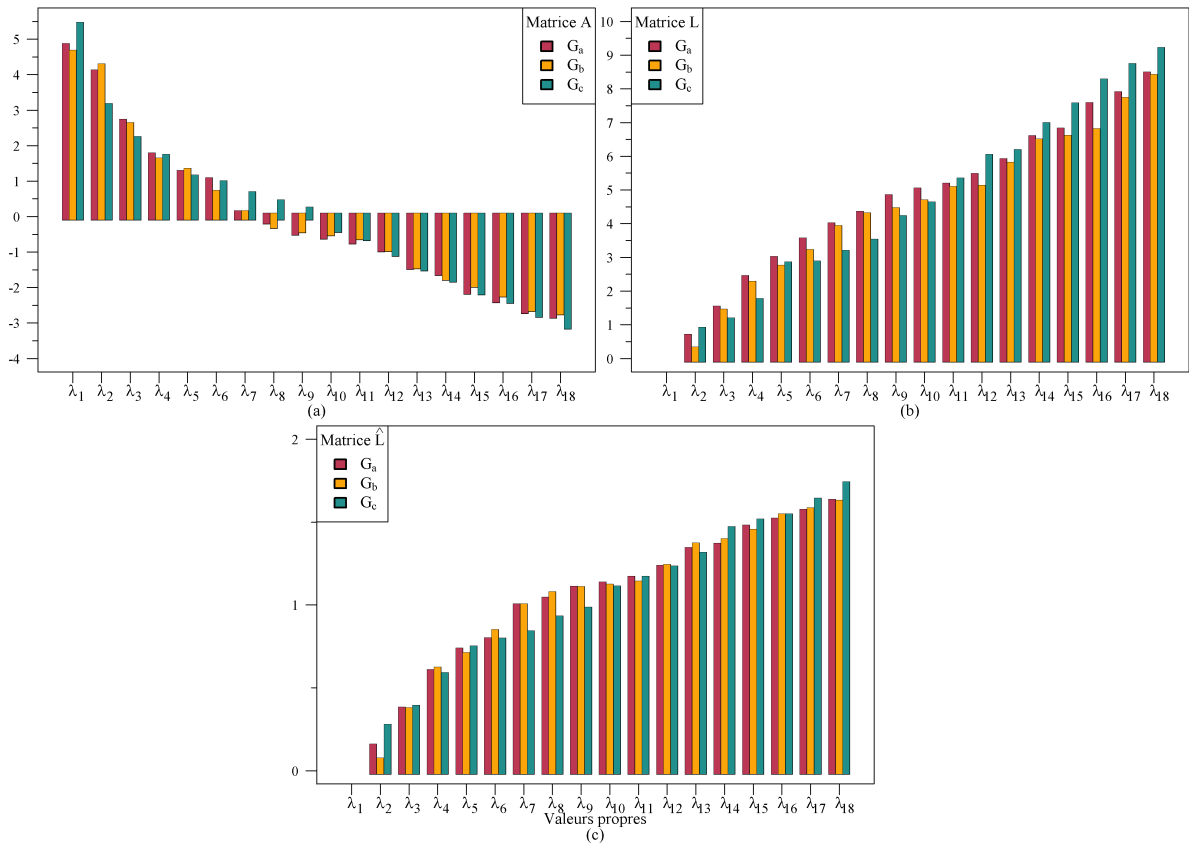


FIGURE 3.4 – Représentation des différents spectres d'amplitudes des valeurs propres définies sur (a) la matrice d'adjacence A , (b) la matrice Laplacienne L , et (c) la matrice Laplacienne normalisée \hat{L} .

Sous cette forme, la distance spectrale entre deux graphes G_1 et G_2 se définit comme la distance entre les valeurs propres de leur matrice d'adjacence respective A_1 et A_2 . Dans ce cas plusieurs types de distances peuvent être choisies comme par exemple ici la distance euclidienne ou celle de Manhattan[126] selon :

$$d_{spec}(G_1, G_2) = \sqrt{\sum_{i=1}^N (\lambda_i^{A_1} - \lambda_i^{A_2})^2} \quad (3.10)$$

$$d_{spec}(G_1, G_2) = \sum_{i=1}^N |\lambda_i^{A_1} - \lambda_i^{A_2}| \quad (3.11)$$

L'utilisation du spectre d'un graphe présente l'avantage d'être par définition invariant à toutes permutations de la matrice d'adjacence. Ainsi, deux graphes isomorphes présentent un spectre de leurs valeurs propres identiques, toutefois la réciproque n'est pas vraie où deux graphes présentant un spectre identique ne sont pas nécessairement isomorphes, c'est le phénomène de cospectralité [125].

Le problème de la cospectralité entre graphes différents entâche l'utilisation de la distance spectrale la rendant peu opérante, d'autant plus que cette cospectralité n'est pas un phénomène rare, notamment concernant les graphes de petits ordres N [127]. De plus, sous cette forme, les valeurs propres de la matrice d'adjacence sont peu informatives sur la topologie du graphe. Afin d'améliorer les performances de la distance spectrale, d'éviter les phénomènes de cospectralité et d'augmenter le caractère informatif des valeurs propres, un ensemble de matrices dérivées de la matrice d'adjacence ont été proposées.

C'est le cas par exemple de la matrice Laplacienne (L) d'un graphe G définie comme :

$$L = D - A \quad (3.12)$$

où D est la matrice diagonale des degrés de A où chaque élément $D(v_i, v_i) = d_{v_i}$. Cette matrice laplacienne est symétrique, et semi définie positive ce qui se traduit par un ensemble de valeurs propres $\lambda_k^L \geq 0$ avec $0 = \lambda_1^L \leq \lambda_2^L \leq \dots \leq \lambda_N^L$ (Fig 3.4b). Cette fois-ci les valeurs propres associées à la matrice Laplacienne rendent compte dans une certaine mesure de l'organisation topologique du graphe et notamment de sa connexité et de l'existence des communautés. En effet, considérant la matrice Laplacienne d'un graphe G , avec toujours $\lambda_1^L = 0$, si $\lambda_2^L > 0$ le graphe est connexe mais si $\lambda_k^L = 0$ et $\lambda_{k+1}^L \neq 0$, alors il existe dans le graphe G , k composantes connexes. Cette interprétation est rendue possible par le fait

que les valeurs propres associées à la matrice Laplacienne rendent compte de la structure "en blocs" de la matrice d'adjacence ou chaque bloc est une composante connexe. Cette propriété peut être relâchée pour identifier les communautés au sein d'un graphes connexe, si $\lambda_k^L \approx 0$ et $\lambda_{k+1}^L \neq 0$ alors le graphe G contient k communautés, dans ce cas, les valeurs propres rendent compte de l'existence de "blocs" faiblement connectés entre eux dans la matrice d'adjacence. Cette propriété peut être observée sur la version modifiée du graphe de Sampson G_b présentant une organisation en deux groupes de noeuds bien distincts.

Finalement, il existe également une forme normalisée de la matrice Laplacienne (\hat{L}) définie selon :

$$\hat{L} = \begin{cases} 1 & \text{if } (v_i = v_j) \\ -\frac{1}{\sqrt{d_{v_i}d_{v_j}}} & \text{if } v_i \text{ et } v_j \text{ sont adjacent} \\ 0 & \text{sinon} \end{cases} \quad (3.13)$$

et permettant de réduire l'influence, s'ils existent, des noeuds présentant un fort degré (hubs) sur la matrice diagonale des degrés D .

Les analyses spectrales sont ainsi des incontournables des approches de recherche et d'identification des communautés au sein d'un graphe. Toutefois, au delà du problème lié au phénomène de cospéctralité, les distances spectrales sont également limitées lorsqu'il s'agit de comparer des graphes d'ordres N différents, puisque cet ordre conditionne le nombre de valeurs propres associées à chaque matrice, d'adjacence ou Laplacienne.

3.2.6 Les "graphlet-based distances"

La dernière catégorie de distances présentée ici est représentée par les "*graphlet-based distances*" qui pourrait être traduit par "distance basée sur les graphlets". Le terme de graphlet a été introduit la première fois en 2004, afin de désigner un ensemble de petits graphes (d'ordre faible) et connexes [128] . Dit autrement, les graphlets représentent toutes les possibilités pour connecter un ensemble de N noeuds (Fig 3.5). Il y a ainsi, une unique façon de connecter deux noeuds (graphlet d'ordre 2, G_0), deux façons de connecter 3 noeuds (graphlet d'ordre 3, G_1 et G_2) et ainsi de suite où le nombre de façons de connecter N noeuds (graphlet d'ordre N) augmente de façon exponentielle avec N . Les distances à base de graphlets reposent toutes sur un socle commun consistant dans un premier temps à identifier dans un graphe G , l'ensemble des graphlets qui le compose.

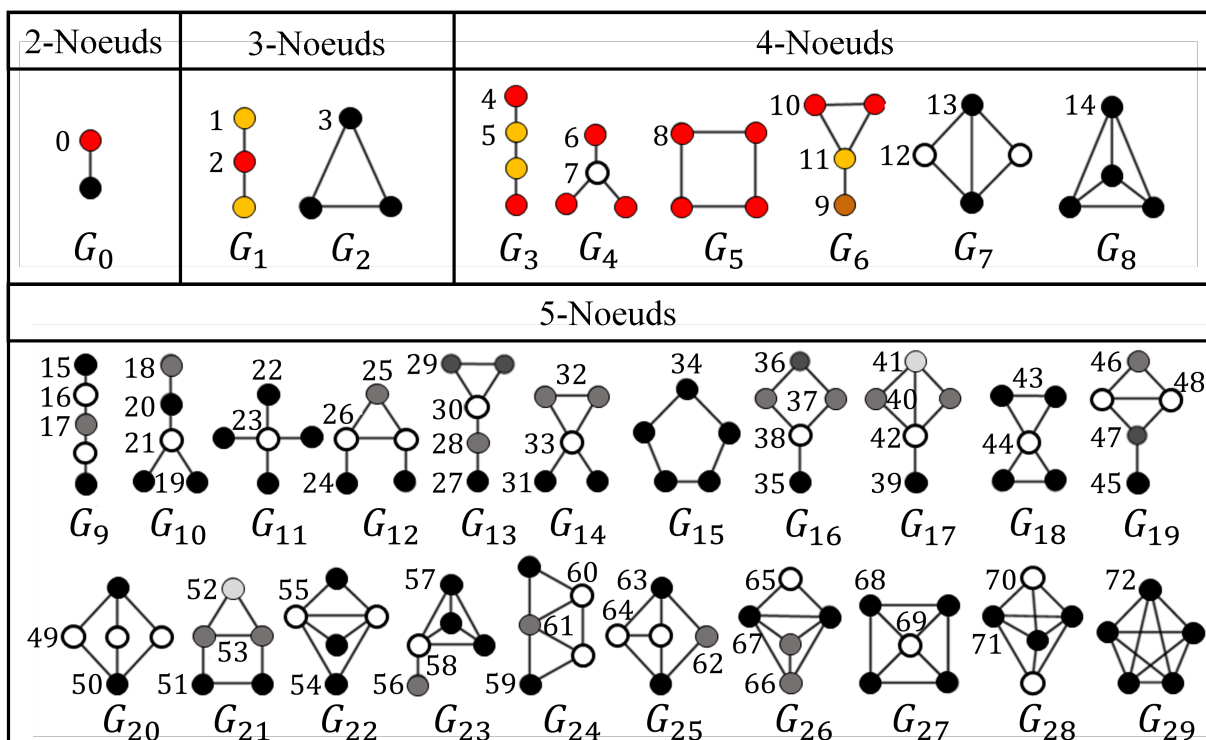


FIGURE 3.5 – Représentation des 30 premiers graphlets d'ordre 2 à 5 et de leur 73 orbites associées (en couleur rouge, jaune et orange, les 11 orbites "non redondantes").

La "Relative Graphlet Frequency distance" (RGF) :

La "relative graphlet fréquence distance" qui peut se traduire par "distance des fréquences relatives des graphlets" est la première distance à base de graphlets parue en 2004 [128]. A partir d'un graphe G , il s'agit d'identifier le nombre d'occurrences de chaque graphlet présentes dans G . Par exemple, soit un graphe G constitué de 3 noeuds formant chaîne, ce graphe est constitué de 3 graphlets, 2 graphlets G_0 et une graphlet G_1 (Fig 3.6).

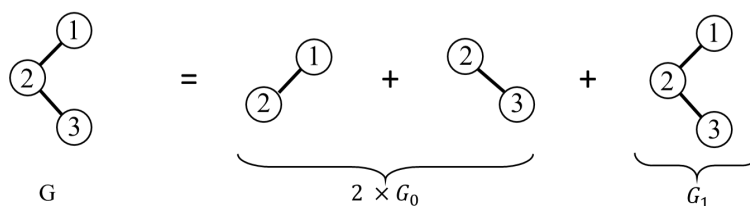


FIGURE 3.6 – Exemple de décomposition d'un graphe G en une somme de graphlets.

La recherche et l'identification des graphlets au sein d'un graphe peut être une opération coûteuse en temps de calcul, d'autant plus lorsqu'il s'agit de les rechercher au sein de grands graphes pouvant contenir plusieurs milliers ou centaines de milliers de noeuds. Ainsi dans cette première distance, seules les 29 graphlets d'ordre 3 à 5 de G_1 à G_{29} sont considérées pour décomposer les graphes.

Dans un graphe G , chaque graphlet G_i avec $i = 1, \dots, 29$ est dénombrée et sa fréquence relative f_i est définie selon :

$$f_i(G) = \frac{n_i(G)}{T(G)} \tag{3.14}$$

avec, $n_i(G)$ le nombre d'occurrences de la graphlet G_i dans G et $T(G)$ le nombre total de graphlets, soit la somme des n_i avec $i = 1, \dots, 29$ (Fig 3.7). Sur la base de ces fréquences relatives, la distance entre deux graphes G_1 et G_2 résumés par leur fréquences relatives f_i^1 et f_i^2 est définie selon :

$$d_{RGF}(G_1, G_2) = \sum_{i=1}^{29} |F_i(G_1) - F_i(G_2)| \tag{3.15}$$

Soit la somme des différences absolues de leurs fréquences des graphlets avec $F_i(G) = -\log(n_i(G)/T(G))$ où la transformation en logarithme doit limiter les effets des graphlets sur-représentées dans un graphe.

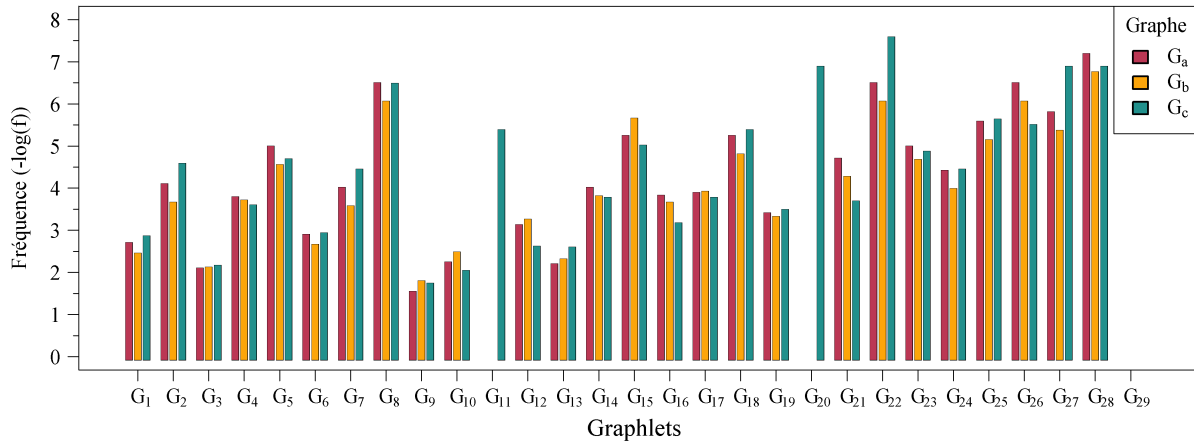


FIGURE 3.7 – Les distributions des fréquences des 29 graphlets de G_1 à G_{29} des trois graphes d'illustrations G_a , G_b et G_c .

La décomposition d'un graphe sous la forme de graphlet est apparue comme un puis-

sant outil permettant d'interroger l'organisation locale (à l'échelle de quelques noeuds) d'un graphe. De plus, sous cette forme de fréquences relatives d'un nombre de graphlet limité, la RGF présente l'énorme avantage d'autoriser la comparaison de graphes de densité d et d'ordre N différents. La RGF apparaît aujourd'hui comme l'approche pionnière des distances à base de graphlets établissant une base au développement et à l'élaboration d'autres distances.

La "Graphlet Degree Distribution Agreement" (GDDA) :

La "*graphlet degree distribution agreement*" qui pourrait être traduit par "concordance des distributions des degrés des graphlets" est un prolongement de la "*relative graphlet frequency distance*" tenant compte d'un plus grand nombre de facettes de l'organisation d'un graphe. A cette fin, une couche supplémentaire est ajoutée à la conceptualisation des graphlets, la notion d'orbite [129]. Chaque graphlet est décomposée en orbites, traduisant l'ensemble des positions non-isomorphes que peut prendre un noeud au sein de cette même graphlet. Par exemple, au sein de l'orbite G_1 décrivant une chaîne à 3 noeuds, un noeud peut soit intervenir dans cette graphlet en position centrale (orbite O_2) ou bien à l'une des deux extrémités de la chaîne (orbite O_1). Ainsi, les 30 graphlets d'ordre 2 à 5 G_i avec $i = 0, \dots, 29$ peuvent être décomposées en 73 orbites notées O_j avec $j = 0, \dots, 72$ (Fig 3.5).

Cette notion d'orbite découle de travaux s'intéressant à l'organisation des graphes d'interactions des protéines où la position qu'occupe un noeud (une protéine) au sein d'un groupe de noeuds, peut aider à renseigner sur sa fonction physiologique [130]. Dit autrement, la notion d'orbite est une réponse à la nécessité de pouvoir interroger le "rôle" qu'occupe un noeud au sein d'un groupe de noeuds.

Après la décomposition en graphlet par la RGF, la GDDA propose de décomposer un graphe en un ensemble d'orbites. La première étape consiste à identifier pour chaque noeud d'un graphe G le nombre de fois où il intervient dans chacune des 73 orbites O_j des 30 graphlets d'ordre 2 à 5 G_i . En reprenant l'exemple précédent du graphe G formant une chaîne à 3 noeuds, les noeuds n°1 et n°3 interviennent chacun une fois dans l'orbite O_0 et dans l'orbite O_1 , tandis que le noeud n°2 intervient deux fois dans l'orbite O_0 et une fois dans l'orbite O_2 (Fig 3.8).

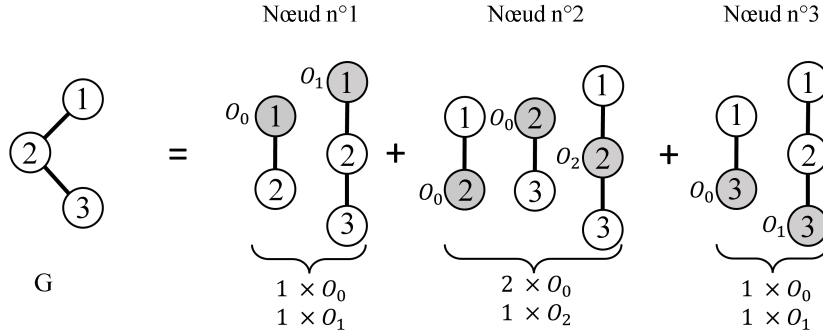


FIGURE 3.8 – Exemple de décomposition d'un graphe G en une somme d'orbites.

L'orbite O_0 correspond au nombre de fois où un noeud v_i est adjacent à un autre noeud v_j , et correspond finalement à la définition de son degré. Sur la base de cette constatation, la décomposition en orbite s'apparente à une généralisation de la notion de degré sous la forme de degré des orbites des graphlets. Ainsi, considérant un graphe G d'ordre N , une matrice A_{GDV} (pour "graphlet degree vector") de dimension $N \times 73$ peut être complétée où chaque ligne i correspond au nombre de fois où le noeud v_i est intervenu dans l'orbite O_j .

$$A_{GDV} = \begin{matrix} & O_0 & O_1 & \cdots & O_j & \cdots & O_J \\ \begin{matrix} v_1 \\ v_2 \\ \vdots \\ v_i \\ \vdots \\ v_{N-1} \\ v_N \end{matrix} & \begin{pmatrix} d_0(v_1) & d_1(v_1) & \cdots & d_j(v_1) & \cdots & d_J(v_1) \\ d_0(v_2) & d_1(v_2) & \cdots & d_j(v_2) & \cdots & d_J(v_2) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ d_0(v_i) & d_1(v_i) & \cdots & d_j(v_i) & \cdots & d_J(v_i) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ d_0(v_{N-1}) & d_1(v_{N-1}) & \cdots & d_j(v_{N-1}) & \cdots & d_J(v_{N-1}) \\ d_0(v_N) & d_1(v_N) & \cdots & d_j(v_N) & \cdots & d_J(v_N) \end{pmatrix} \end{matrix} \quad (3.16)$$

Chaque colonne j de la matrice contient le vecteur des degrés de l'orbite O_j permettant de généraliser là aussi, la notion de distribution des degrés, en distribution des degrés d'orbites. Soit un graphe G et une orbite O_j , sa distribution des degrés d'orbite est notée $d_G^j(k)$, soit, dans le graphe G , le nombre de noeuds intervenant k fois dans l'orbite O_j .

Cette distribution des degrés est ensuite réduite selon :

$$S_G^j(k) = \frac{d_G^j(k)}{k} \quad (3.17)$$

afin de limiter la contribution des larges degrés, puis "normalisée" selon :

$$N_G^j(k) = \frac{S_G^j(k)}{T_G^j} \quad \text{Avec} \quad T_G^j = \sum_{k=1}^{\infty} S_G^j(k) \quad (3.18)$$

La GDDA repose alors sur une comparaison entre les distributions "normalisée" des degrés des différentes orbites O_j de deux graphes G_1 et G_2 , dans un premier temps sous la forme d'une distance :

$$d^j(G_1, G_2) = \frac{1}{\sqrt{2}} \left(\sum_{k=1}^{\infty} (N_{G_1}^j(k) - N_{G_2}^j(k))^2 \right)^{\frac{1}{2}} \quad (3.19)$$

avec $d^j(G_1, G_2) = 0$ les deux distributions sont identiques et $d^j(G_1, G_2) = 1$ les deux distributions sont très différentes. Le terme $1/\sqrt{2}$ est une constante de normalisation assurant une distance bornée $0 < d^j(G_1, G_2) < 1$ [131]. Puis cette distance est renversée sous la forme d'une similarité (*agreement* \approx concordance), selon :

$$A^j(G_1, G_2) = 1 - d^j(G_1, G_2), \quad \text{pour } j \in \{0, 1, \dots, 72\} \quad (3.20)$$

Où finalement la *GDDA* entre deux graphes est définie comme la moyenne des similarités de leurs orbites :

$$GDDA(G_1, G_2) = \frac{1}{73} \sum_{j=0}^{72} A^j(G_1, G_2) \quad (3.21)$$

En décomposant profondément l'organisation d'un graphe au travers de ses orbites, la GDDA s'avère particulièrement efficace pour évaluer la similarité entre graphes de même ordre N et de même densité d . Toutefois la GDDA a suscité de vives critiques la jugeant "instable" lors de la comparaison de graphes de faibles ordres, requalifiée de "sensible" par ses auteurs [131]. De plus, en reposant sur une comparaison entre des distributions des degrés, la GDDA présente les mêmes sensibilités à l'ordre et à la densité que celles mis en évidence dans les "*features-based distances*", la rendant peu opérante aux applications nécessitant la comparaison de graphes d'ordres et ou de densités différentes.

La "Graphlet Correlation Distance" (GCD) :

De la même manière que la GDDA est un prolongement de la RGF, la "*graphlet correlation distance*" qui pourrait être traduit par "distance des corrélations des graphlets"

est un prolongement de la GDDA proposé par Yaveroğlu et al en 2014 [132]. Cette dernière distance devant répondre aux limites de la précédente en partage les fondements mais se distingue très vite sous la forme d'une distance originale, laissant de côté les distributions des orbites pour se concentrer sur l'analyse de la matrice des vecteurs des degrés A_{GDV} .

La GCD repose sur l'idée qu'un graphe décrit au travers des degrés des orbites des graphlets, peut être résumé par les corrélations qui existent entre ses orbites O_j . Ainsi, soit un graphe G et sa matrice des degrés des 73 orbites des graphlets A_{GDV} , la GCD repose sur la construction d'une matrice intermédiaire, appelée matrice des corrélations des graphlets, notée GCM , contenant l'ensemble des coefficients de corrélation de Spearman [133] entre chaque paire d'orbites O_i, O_j avec $i, j = 0, \dots, 73$, selon :

$$GCM_G(i, j) = cor_{spearman}(A_{GDV}(:, i), A_{GDV}(:, j)) \text{ avec } i, j = 0, \dots, 73 \quad (3.22)$$

Sous cette forme, la GCM est une matrice symétrique de dimension 73×73 contenant les coefficients de corrélation de Spearman calculés entre chaque paire de colonnes de la matrice A_{GDV} . L'utilisation des 73 orbites est un reliquat de la GDDA où Yaveroğlu et al ont finalement montré que l'utilisation d'un nombre plus limité de 11 orbites, qualifiées de non-redondantes (Fig 3.5) était plus pertinente que l'utilisation des 73 orbites proposées pour la GDDA. Ainsi, la GCM d'un graphe G se voit réduite sous la forme d'une matrice 11×11 , permettant de résumer la topologie d'un graphe.

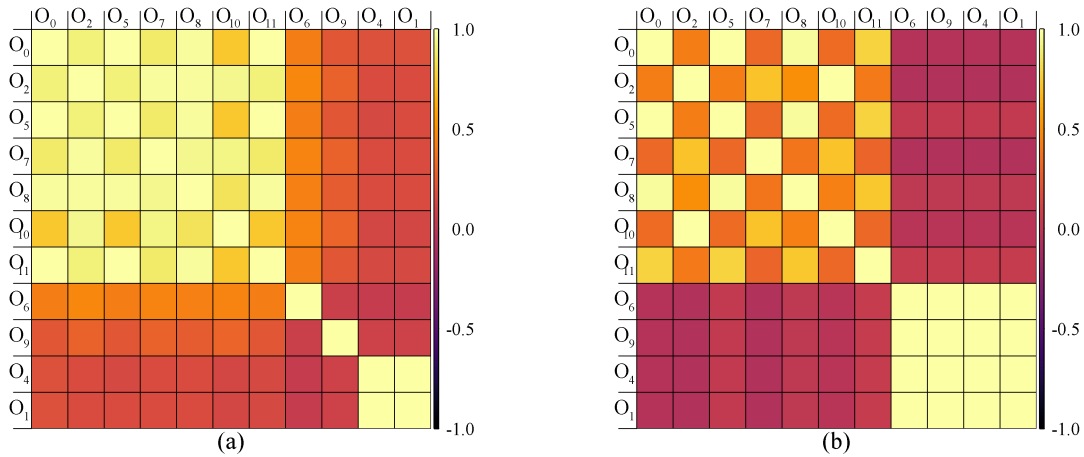


FIGURE 3.9 – Illustration de deux matrices de corrélation des graphlets (a) d'un graphe d'Erdős-Rényi $G(1000, 0.01$ et (b) d'un graphe scale-free Barabasi-Albert $G_{BA}(1000, 2)$

En effet, les auteurs ont montré que les différentes corrélations pouvaient renseigner sur l'organisation topologique d'un graphe. Ainsi par exemple, une forte corrélation entre

les orbites O_0, O_2, O_5 et O_7 est caractéristique de l'existence de hubs [132] (Fig 3.9b), de même concernant les orbites O_{10} et O_{11} traduisant l'existence de communautés au sein du graphe, et ainsi de suite. La GCM fournit ainsi un résumé de la topologie globale d'un graphe sur la base d'une analyse locale effectuée au travers des graphlets.

Finalement la distance GCD entre deux graphes G_1 et G_2 , respectivement résumé au travers de leurs matrices des corrélations des graphlets CGM_{G_1} et CGM_{G_2} , est définie comme la distance euclidienne les séparant :

$$GCD_{11}(G_1, G_2) = \sqrt{\sum_{i=1}^{11} \sum_{j=i+1}^{11} (GCM_{G_1}(i, j) - GCM_{G_2}(i, j))^2} \quad (3.23)$$

dénotée GCD_{11} en référence aux 11 orbites sélectionnées pour construire les GCM.

La GCD est ainsi une distance extrêmement modulable quant au choix des orbites O_j à considérer pour réaliser la comparaison, au delà de la GCD_{11} il existe par exemple la GCD_{73} . De plus en résumant l'information sous la forme d'une matrice de dimension invariante (dépendante du nombre d'orbites considérée), la GCD autorise la comparaison de graphes d'ordre ou de taille différents. En effet, d'après les auteurs, en reposant sur la corrélation entre paire d'orbites, la GCM est présentée comme peu sensible aux différences d'ordre N et/ou de densité d entre les graphes à comparer.

Bilan, quelle distance choisir ?

Nous l'avons vu, il existe une multitude de méthodes permettant de comparer des graphes, chacune présentant son lot d'avantages et/ou d'inconvénients. Le choix d'une méthode plus qu'une autre est ainsi conditionné d'une part au regard de l'application qui lui est réservée et de sa flexibilité à réaliser cette dernière, et d'autre part, par ses performances effectives conditionnant la confiance qu'on est en mesure d'accorder à ses résultats. Une "bonne" distance est une distance capable d'attribuer une distance "faible" entre des graphes similaires et une distance élevée entre des graphes différents.

Sur ce point, l'ensemble des distances présentées ici remplit cet objectif au regard de la comparaison illustrative réalisée sur les deux versions du graphe de Sampson et du graphe d'Erdős-Rényi (Table 3.1). En effet, toutes ces distances ont attribué une plus faible distance (ou une plus forte similarité dans le cas de la GDDA) entre les deux versions très similaires du graphe de Sampson par rapport à celle entre le graphe de Sampson et sa version d'Erdős-Rényi.

TABLE 3.1 – Comparaison des résultats des méthodes de comparaison entre deux versions du graphe de Samspon et d'un graphe d'Erdős-Renyi.

Comparaison	Features	Spectrales			Graphlet		
	NetSimile	A	L	\hat{L}	RGF	GDDA	GCD 11
G_a vs G_b	0.041	0.604	1.221	0.125	7.762	0.879	1.275
G_a vs G_c	5.989	1.616	2.420	0.315	21.269	0.705	3.195

Ce n'est évidemment pas au travers de cet exemple simpliste, que les contours des performances de ces différentes méthodes peuvent être esquissées afin de choisir la "meilleure" méthode. Cependant dans la littérature, certain travaux ont essayé d'esquisser ces contours [112, 113, 134] et ont révélé que les distances à base de graphlets et notamment la GCD présentaient les meilleurs comportements et les meilleures performances établis au travers d'épreuves de classifications d'ensemble de graphes. Nous nous appuyons ainsi sur cette littérature pour choisir la GCD et plus précisément la GCD_{11} comme méthode d'analyse des graphes de flottille.

3.3 La comparaison de graphes, un problème de taille

A ce stade, nous disposons d'un côté de graphes mettant en relation des ensembles de trajectoires au travers de leurs relations dyadiques, de l'autre, des modèles de graphes aléatoires, dont le modèle d'Erdős-Rényi incarnant l'indépendance des arêtes, et d'une distance permettant de comparer des graphes, la GCD_{11} .

Dans cette section, il doit être question d'une part, de répondre de la manière dont doivent être construits les graphes empiriques afin de pouvoir les comparer aux réalisations du modèle d'Erdos-Renyi, et d'autre part, de statuer sur la manière dont ces graphes doivent être comparés afin que le résultat de cette comparaison permette de valider ou non, l'hypothèse d'indépendance.

3.3.1 Vers une réduction spatiale et temporelle des graphes de flottille

Deux trajectoires ayant eu lieu à des moments différents, par exemple à une semaine d'intervalle, se traduisent dans un graphe sous la forme d'un couple de noeuds ayant une probabilité nulle d'être reliés par une arête. De même, deux trajectoires relevant de navires

inféodés à des ports très éloignés l'un de l'autre ont une probabilité faible, voire nulle de se rencontrer. Ainsi, sans être capable d'estimer précisément les probabilités que deux trajectoires se rencontrent en fonction de leur "moment" et de leur "lieu" de réalisation, le simple fait de pouvoir affirmer que ces probabilités ne sont pas uniformes entre tous les couples de trajectoires, va à l'encontre des conditions du modèle d'Erdős-Rényi, et est suffisant à la réfutation de l'hypothèse d'indépendance de leurs relations dyadiques.

Afin de rendre la comparaison entre les graphes empiriques et des réalisations du modèle d'Erdős-Rényi pertinente, il est nécessaire de limiter les effets de la variabilité spatiale et temporelle des trajectoires. La solution apportée à ce problème est une réduction spatiale et temporelle des graphes de flottille où il s'agit d'isoler des ensembles suffisamment importants de trajectoires s'établissant, d'une part, sur une même "zone spatiale", et d'autre part, dans un même "espace temporel".

Réduction de la flottille OTB :

Du point de vue de la flottille des OTB, cette réduction spatiale et temporelle est réalisée en limitant le cas d'étude, aux trajectoires du seul port de Boulogne-sur-Mer. Plusieurs raisons sous-tendent ce choix, tout d'abord, le port de Boulogne-sur-Mer est le plus important de la Manche à la fois en terme d'effectif de navires et de marées OTB, permettant de fournir un matériel suffisant à la construction de graphes (nombre de noeuds). De plus, l'activité des navires liés à ce port présente une périodicité hebdomadaire (peut être unique dans la Manche) se traduisant par une très forte synchronisation temporelle des marées (Fig 3.10g), débutant le lundi et se terminant en fin de semaine le vendredi ou le samedi, soit des marées d'environ six jours. Cette périodicité spécifique au port de Boulogne-sur-Mer est identifiable au regard de l'histogramme de la distribution des durées des marées OTB à l'échelle de la Manche (Fig 2.5a) présentant un pic entre 140 et 180 heures, soit environ 6 à 7 jours de marées. Toutefois, certains navires réalisent tout de même une succession de deux marées d'environ deux jours et demie chacune.

Ainsi, les trajectoires de la flottille OTB du port de Boulogne-sur-Mer sont analysées semaine après semaine (Fig 3.10d-f) permettant de produire des graphes hebdomadaires (Fig 3.10a-c), théoriquement au nombre de 52 chaque année. En isolant de la sorte des ensembles de trajectoires partageant le même espace temporel et spatial, l'hypothèse peut être faite que tous les couples de trajectoires ont une probabilité similaire d'entretenir une relation dyadique et par conséquent, que l'organisation des graphes qui en découle n'est

représentative que du comportement collectif des pêcheurs.

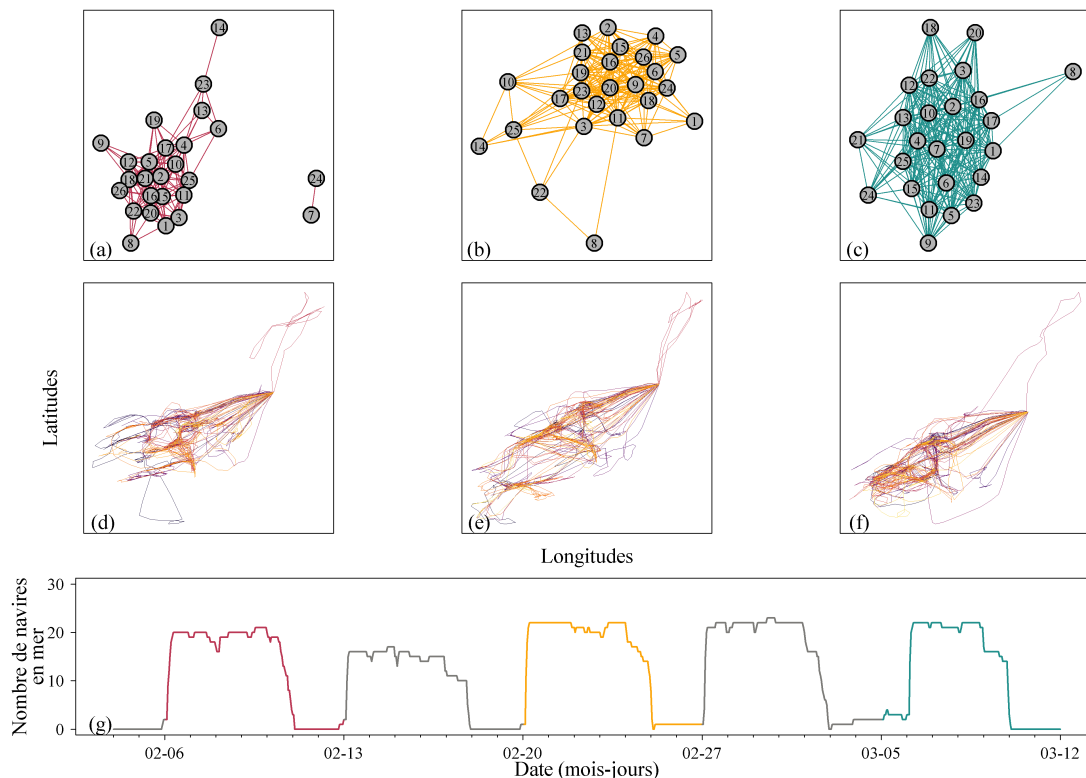


FIGURE 3.10 – L'activité périodique des navires de la flottille OTB du port de Boulogne-sur-Mer (g) son découpage hebdomadaire (d-f) et leur graphe associé (a-c).

Réduction de la flottille PTM :

Concernant la flottille des PTM, la situation est bien plus délicate (Fig 3.11). Tout d'abord au regard de l'effectif des navires qui constituent cette flottille, déjà faible à l'échelle de la Manche entière, qui s'en retrouve encore plus réduit lorsqu'il s'agit de considérer un unique port dont celui de Cherbourg est le plus imposant. D'autre part, contrairement à l'activité des OTB de Boulogne-sur-Mer qui s'étend toute l'année, l'activité des PTM est saisonnière, principalement hivernale, et n'est pas marquée d'une quelconque périodicité hebdomadaire (Fig 3.11g). Dans ce cas, le regroupement des marées en semaines induit un découpage brutal des marées sans lien avec l'activité du port de Cherbourg. Du fait de la saisonnalité de l'activité des PTM, un faible nombre de graphes peut être construit.

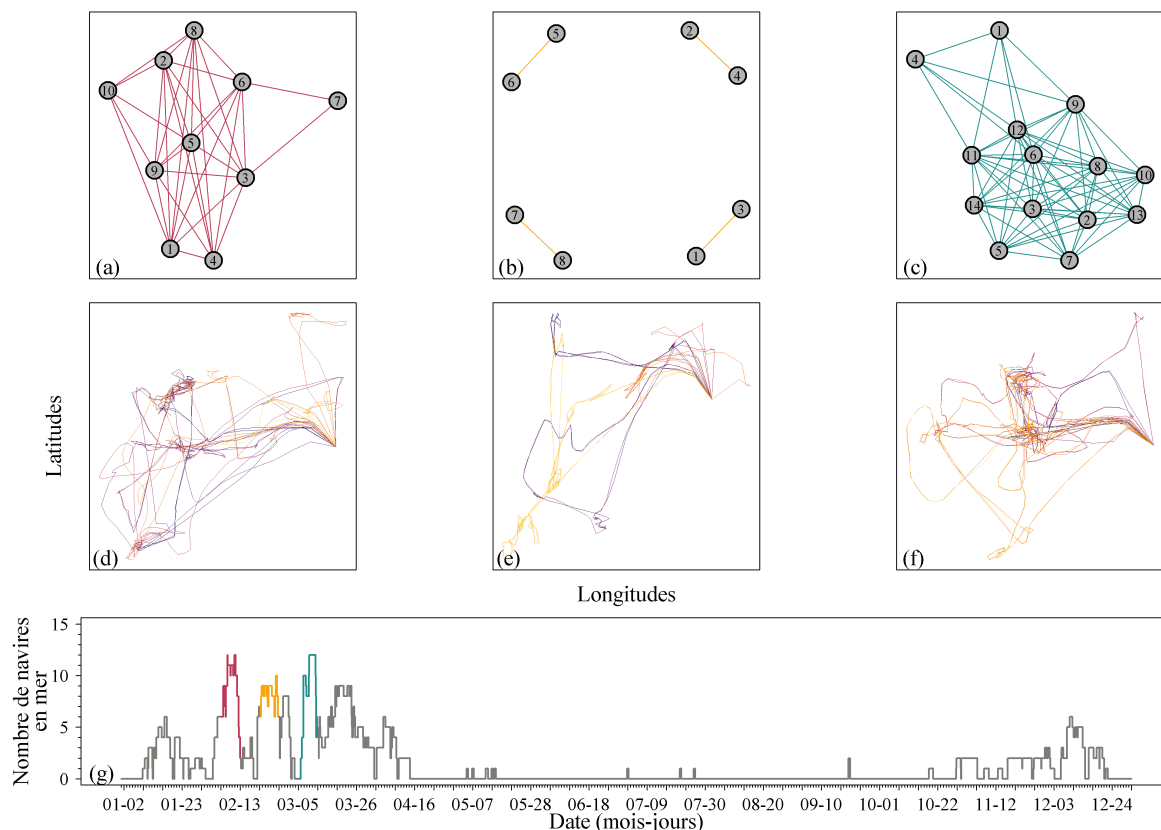


FIGURE 3.11 – L'activité saisonnière non périodique des navires de la flottille PTM du port de Cherbourg (g) son découpage hebdomadaire (d-f) et leur graphe associé (a-c).

La réduction de l'ordre, le contrecoup de la réduction des graphes :

Le contrecoup de la réduction spatiale et temporelle du domaine d'étude à l'échelle de l'activité hebdomadaire d'un port, est une diminution du nombre de trajectoires à intégrer sous la forme d'un graphe se traduisant par des ordres faibles ($N < 52$) (Table 3.2).

TABLE 3.2 – Caractéristiques des graphes hebdomadaires des flottilles PTM de Cherbourg et OTB de Boulogne-sur-mer.

Engin	Année	Ordre N			Densité d	
		Ordre > 10	Étendue	Moyenne	Étendue	Moyenne
PTM	2012	2	[10;14]	12	[0.78;0.8]	0.79
	2013	6	[10;18]	12	[0.22;0.86]	0.55
OTB	2012	47	[11;52]	24	[0.12;0.8]	0.39
	2013	47	[10;50]	22	[0.13;0.87]	0.37

La notion de comportement collectif n'existe, par définition qu'en présence d'un nombre suffisant d'acteurs, il est ainsi dans un premier temps décider de ne conserver que les graphes présentant un ordre $N > 10$. La flottille des PTM est alors réduite à l'expression de 8 graphes (8 semaines d'activité) de très faible ordre $N < 20$ mais de forte densité $d > 0.5$. La faible représentation des graphes de cette flottille est toutefois peu handicapante pour le développement de ce travail du fait de la connaissance a priori du comportement non-indépendant des navires qui la constitue. Le véritable cas d'étude de ce travail, sont en effet les graphes de la flottille des OTB au comportement collectif inconnu, qui quant à elle fournit une large collection de près d'une centaine de graphes d'ordres plus importants pouvant dépasser les 50 noeuds.

Malgré tout, les graphes de la flottille des OTB, du haut de leur 52 noeuds au maximum, bien qu'entrant en résonance avec les graphes du siècle dernier, apparaissent comme des anomalies dans le monde des graphes modernes, représentés par des graphes dont l'ordre dépasse aisément les centaines ou les milliers de noeuds. Cette considération est à mettre en regard avec la littérature évoquée précédemment s'intéressant aux performances des méthodes de comparaison de graphes qui s'inscrivent dans le cadre des caractéristiques des graphes couramment rencontrés, c'est-à-dire précisément, l'opposée des graphes de flottilles, à savoir, des graphes d'ordre élevé ($N \gg 100$) et de faible densité $d < 0.1$.

Les performances de la *GCD* n'ont ainsi été évaluées que dans le monde des grands graphes, où il nous incombe la tâche de plonger la *GCD* dans le monde des petits graphes afin de vérifier qu'elle conserve ses performances, et le cas échéant, de l'inscrire dans une approche permettant de statistiquement statuer sur la similarité entre des réalisations de modèles de graphes et des graphes empiriques, autrement dit, de l'inclure dans un test statistique. Ces deux dernières considérations ont fait l'objet d'un article scientifique paru le 15 février 2023 et représentant la sous-section suivante.

PLOS ONE

RESEARCH ARTICLE

Graphlet correlation distance to compare small graphs

Jérôme Roux^{1*}, Nicolas Bez², Paul Rochet³, Rocío Joo⁴, Stéphanie Mahévas¹

¹ UMR DECOD, IFREMER, BP 21105, Nantes Cedex, France, ² MARBEC, IRD, Univ Montpellier, Ifremer, CNRS, INRAE, Sète, France, ³ ENAC, Toulouse, France, ⁴ Global Fishing Watch, Washington, DC, United States of America

FIGURE 3.12

3.3.2 Vers la comparaison des petits graphes

Introduction

Nous tronquons une partie de l'introduction dont le contenu a déjà été évoqué au cours des précédents chapitre. Le lecteur pardonnera je l'espère le passage à l'anglais.

[...] This work deals with two main gaps in the literature. First, we assess the ability of GCD_{11} to correctly distinguish small simulated graphs from known model types (Erdős-Rényi[135], Fitness Scale-Free[136], Watts-Strogatz small-world[91] and geometric[137]) by a clustering approach[132, 138] using numerical experiments. We simulated graphs of different order and density, fluctuating the order from 10 to 100 to mimic the range encountered in some real small graphs[139-141], while the density range is completely covered from 0 to 1. We study the power of the GCD_{11} to discriminate graphs with equal density and order and graphs that differ in density and order. Then, we perform a sensitivity analysis of the GCD_{11} discrimination power to the values of density and order. This allows defining a domain of separability, i.e. an envelope of orders and densities for which the GCD_{11} allows distinguishing between graphs. Second, we propose a statistical test based on the GCD_{11} to evaluate whether an empirical graph can be considered as an outcome of a particular random graph. Finally, we illustrate the use of the GCD_{11} to analyse small graphs characterising pairwise links between fishing vessels at sea. We test if these graphs can be outcomes of an Erdős-Rényi random graph model which would be indicative of vessels that cross independently at sea, an implicit assumption frequently made in fisheries.

Materials and methods

Graphlet Correlation Distance (GCD_{11})

Yaveroğlu et al [132] recently proposed to compare graphs on the basis of the first eleven non-redundant orbits graphlets of up to four nodes. Considering a graph G of order N , they first compute the $N \times 11$ matrix which contains for each node their orbits' degree i.e. the number of times the node is presented in each of the eleven orbits. The columns are called Graphlet Degree Distributions (GDD)[129] and the first column is the standard vector of degree values. Then, the Spearman's Correlation coefficient[133]

is computed between all columns of the GDD matrix to build an 11×11 matrix called the Graphlet Correlation Matrix (GCM). In this framework, the topology of a given graph G is summarised by its Graphlet Correlation Matrix denoted GCM_G . The GCD_{11} between two graphs G_1 and G_2 is then defined as the Euclidean distance between the upper triangular parts of their respective GCM :

$$GCD_{11}(G_1, G_2) = \sqrt{\sum_{i=1}^{11} \sum_{j=i+1}^{11} (GCM_{G_1}(i, j) - GCM_{G_2}(i, j))^2} \quad (3.24)$$

GCD_{11} on small synthetic graphs

The performance of the GCD_{11} to identify similarities between small graphs is assessed with a numerical experiment using four different models of random graphs, namely the Erdős-Rényi (ER) [135], the Fitness Scale-Free (SF) [136], the Watts-Strogatz small-world (SW) [91] and the Geometric (GO) [137] models.

The Erdős-Rényi model is the simplest and most common uncorrelated random graph model. An Erdős-Rényi graph $ER(N, d)$ of order N and edge density $d = 2m / (N(N - 1))$ gets m edges that are randomly and uniformly chosen among the $\binom{N}{2}$ possible edges [135]. This simple configuration results in an uncorrelated graph i.e., with a zero assortativity [142] meaning that there is no preferential attachment among nodes. In other words, the Erdős-Rényi random model generates graphs where edges are statistically independent of each other.

Fitness Scale-Free models $SF(N, d, \gamma)$ are derived [143] from Scale-Free Barabási-Albert models [144]. A graph is deemed scale-free when its node degree distribution $p(k)$ follows a power law $p(k) \sim k^{-\gamma}$ whose power γ is generally observed empirically between 2 and 3 [145]. As a compromise, we set $\gamma = 2.5$ in this study. In scale-free graphs, few nodes have high connectivity with respect to the average degree leading to the emergence of hubs as observed (or supposedly observed [146]) in some real-world graphs [93]. Fitness models are based on the attractiveness of each node [136]. When this follows a power law, the degree distribution also follows a power law as a scale-free graph [147].

The Watts-Strogatz small-world model was developed to address the lack of realism of the random graph model (Erdős-Rényi) by relying on the notion of "small-world" phenomenon [90] also known as six degrees of separation [148]. The Watts-Strogatz small-world model is an intermediate model between a regular graph [92] and an Erdős-Rényi graph

allowing to reconcile local properties of one and global properties of the other [149]. A Watts-Strogatz small-world graph $SW(N, k, p)$ is generated from a ring lattice graph of order N , where each node is connected to its $2k$ nearest neighbours. Then, each edge is kept with probability $1 - p$ or reconnected to a randomly chosen node (without loop or multiple edges) with a probability p . The parameter p thus controls the coexistence of “short-range” (from the ring lattice) and “long-range” (from the random reconnection) connections [149]. As the edge density is $d = \frac{2k}{N-1}$, it does not span all the possible values on the $[0, 1]$ interval. In the numerical experiment below, we randomly added or removed some edges (between 10 and 50 depending on the order N) to reach the desired edge density when necessary. Some of the simulated graphs were thus no longer Watts-Strogatz small-world strictly speaking but remained highly similar to it. To bring out the small world property [150], we set $p = 0.05$.

The Geometric model allows to generate spatial graphs where the condition of existence of an edge between two nodes depends on their proximity [151]. A Geometric graph $GO(N, l, r)$ is generated by placing N independent nodes uniformly at random in \mathbb{R}^l and by connecting pairs whose distance is smaller than r [137]. In this study, the distance threshold r was chosen to obtain the desired density d . Geometric graphs are dominated by local interactions [152] leading to the emergence of community structure.

For each model $M \in \{ER, SF, SW, GO\}$ and for a given order N and a given edge density d we generate 100 graphs $G_M^i(N, d)$ with $i = 1, \dots, 100$. For $M \in \{ER, SF, SW\}$ we use the R 4.1.3[153] package `igraph` (v1.2.1)[154] to generate graphs (see Data Availability statement).

Comparing graphs with the same order and edge density

We define a set of 51 possible values of order N describing two gradual incremental steps as $N \in \{10, 11, \dots, 48, 49, 50, 55, 60, \dots, 90, 95, 100\}$ and a set of 101 possible values of edge density d regularly spaced as $d \in \{0, 0.01, \dots, 0.99, 1\}$. According to these two sets of order and density, for a given order N , and a given edge density d , for each couple of models $(M_1, M_2) \in \{ER, SF, SW, GO\}^2$ with $M_1 \neq M_2$, we computed all the pairwise GCD_{11} between their 100 respective generated graphs to construct a 200×200 block symmetric distance matrix $D = \begin{bmatrix} D_{M_1, M_1} & D_{M_1, M_2} \\ D_{M_2, M_1} & D_{M_2, M_2} \end{bmatrix}$. The discriminating power of GCD_{11} is assessed by the Area Under the Precision-Recall (AUPR) curve[138] computed as follows. For each

distance threshold $\epsilon_k, k = 1, \dots, 100$, regularly spanning the range of distance values of D , we compute :

- the true positives TP , as the number of distances between graphs from the same model smaller than ϵ_k ;
- the true negatives TN , as the number of distances between graphs from two different models greater or equal to ϵ_k ;
- the false negatives FN , as the number of distances between graphs from the same model greater or equal to ϵ_k ;
- and the false positives, FP , as the number of distances between graphs from two different models smaller than ϵ_k .

Precision (P) and recall (R) are then defined as :

$$P(\epsilon) = \frac{TP(\epsilon)}{TP(\epsilon)+FP(\epsilon)} \quad (3.25)$$

$$R(\epsilon) = \frac{TP(\epsilon)}{TP(\epsilon)+FN(\epsilon)} \quad (3.26)$$

The diagonals of D_{M_1, M_1} and D_{M_2, M_2} are trivial (null distance between a graph and itself) and are thus excluded from these calculations. We also removed the diagonals of D_{M_2, M_1} and D_{M_1, M_2} when computing the area under the precision-recall curve to guarantee the same number of intra-model distances D_{M_1, M_1} and D_{M_2, M_2} and inter-model distances D_{M_2, M_1} and D_{M_1, M_2} . The area under the curve of the precision given the recall curve (AUPR) is defined as :

$$AUPR_{M_1, M_2}(N, d) = \sum_{k=2}^{100} P(\epsilon_k) \Delta R(\epsilon_k) \quad (3.27)$$

where $\Delta R(\epsilon_k)$ is the change in recall from rank $k - 1$ to k . For each combination of order N and edge density d , the resultant $AUPR_{M_1, M_2}(N, d)$ is used to fill an $|N| \times |d|$ matrix A_{M_1, M_2} . The discriminating power of the GCD_{11} amongst the four tested models is set to the minimum value obtained over all possible comparisons. From the $\binom{4}{2}$ pairs of models, we thus finally fill a $|N| \times |d|$ matrix A as :

$$A(N, d) = \min_{i \neq j} (A_{M_i, M_j}(N, d)) \quad (3.28)$$

An AUPR score of 1 means a perfect distinction between M_1 and M_2 (i.e. two clusters without overlapping) whereas an AUPR score of 0.5 is the expected score of a random classifier. An AUPR score of 0 occurs when graph topologies are all identical. We arbi-

trarily consider that an AUPR larger than 0.9 allows discrimination between two models amongst the four that are used and defined the domain of separability as the set of orders and densities providing an AUPR larger than 0.9. In the domain of separability, the GCD_{11} is small (resp. large) between graphs coming from the same (resp. different) models amongst the four models used in the study. In other words, the domain of separability reflects the ability of the GCD_{11} to discriminate graphs of two different models into two weakly or non-overlapping clusters.

Comparing graphs with different orders and edge densities

For each couple $(M_1, M_2) \in \{ER, SF, SW, GO\}^2$ with $M_1 \neq M_2$, and for all possible pairs of combinations of orders and densities $(N, d) \times (N', d')$ we also build the three following 100×100 GCD_{11} matrices using the already simulated graphs :

$$D_{M_1(N,d),M_1(N',d')}[i, j] = GCD_{11}(G_{M_1}^i(N, d), G_{M_1}^j(N', d')) \quad (3.29)$$

$$D_{M_2(N,d),M_2(N',d')}[i, j] = GCD_{11}(G_{M_2}^i(N, d), G_{M_2}^j(N', d')) \quad (3.30)$$

$$D_{M_1(N,d),M_2(N',d')}[i, j] = GCD_{11}(G_{M_1}^i(N, d), G_{M_2}^j(N', d')) \quad (3.31)$$

for $i, j = 1, \dots, 100$. We then computed the proportion of cases where the inter-model distance $D_{M_1(N,d),M_2(N',d')}$ was larger than the two intra-model distances $D_{M_1(N,d),M_1(N',d')}$ and $D_{M_2(N,d),M_2(N',d')}$. We consider this proportion as the probability that the GCD_{11} assigns a smaller distance between two graphs from the same model than between two graphs from different models. A probability of 1 means that all intra-model distances are smaller than all inter-model distances. This probability is used to fill an $(|N| \times |d|) \times (|N'| \times |d'|)$ asymmetric matrix of probability $B_{M_1, M_2}((N, d), (N', d'))$. As in the previous section, all the possible comparisons were combined into a $(|N| \times |d|) \times (|N'| \times |d'|)$ synthetic matrix of probability as :

$$B((N, d), (N', d')) = \min_{k \neq l} \left(B_{M_k, M_l}((N, d), (N', d')) \right) \quad (3.32)$$

We arbitrarily consider that a probability larger than 0.9 allows discrimination between two models amongst the four that are used and allows us to define the domain of separability. We also defined the surface of the domain of separability as the proportion of cells in the matrix of probability where the value is larger than 0.9. To limit the computing time and because the outputs change slowly with the order values, the number of possible values

for the order and density are reduced so that $(N, d) \in \{10, 20, \dots, 100\} \times \{0, 0.02, \dots, 1\}$.

To visualize distances between graphs we use a principal component analysis (PCA) [111] where each observation (graph) is described by 55 variables (the 55 graphlet correlation coefficients in the GCM). Since PCA is based on the Euclidean distance as the GCD_{11} , to preserve the properties of the GCD , the PCA is computed on the 55 centered variables without scaling.

Statistical test

In order to test if an empirical graph $G(N, d)$ is an outcome of an $M(N, d)$ random graph model (H_0) with $M \in \{ER, SF, SW, GO\}$, we build the following randomization statistical test. First, we simulate 1 000 independent outcomes M_k ($k = 1, \dots, K = 1\,000$) of each possible reference model M . Second, we compute their Graphlet Correlation Matrices $GCM(M_k)$ and their average :

$$\overline{GCM}_M = \frac{1}{K} \sum_{k=1}^K GCM(M_k) \quad (3.33)$$

where \overline{GCM}_M denotes the average Graphlet Correlation Matrix of M . Third, we compute the Graphlet Correlation Distance δ_{M_k} between $GCM(M_k)$ and \overline{GCM}_M and the Graphlet Correlation Distance δ_G between GCM_G and \overline{GCM}_M :

$$\delta_{M_k} = \sqrt{\sum_{i=1}^{11} \sum_{j=i+1}^{11} \left(\overline{GCM}_M(i, j) - GCM(M_k)(i, j) \right)^2} \quad (3.34)$$

$$\delta_G = \sqrt{\sum_{i=1}^{11} \sum_{j=i+1}^{11} \left(\overline{GCM}_M(i, j) - GCM_G(i, j) \right)^2} \quad (3.35)$$

Under H_0 , $P(\delta_G < \delta) = P(\delta_{M_k} < \delta)$ with $\delta \in \mathbb{R}^+$, and the p -value for testing H_0 is calculated as $P(\delta_{M_k} > \delta_G)$. We computed η as the number of times the distance δ_G between GCM_G and \overline{GCM}_M is smaller or equal than the distance δ_{M_k} . The p -value is then defined by $\hat{p} = (\eta + 1)/(K + 1)$ [155]; the larger the p -value, the less evidence against H_0 . To account for the difference in variability between the correlation coefficients of each pair of orbits, we also investigated the use of a standardised distance which provided very

similar outcomes (S1 Eq).

Empirical graphs

The developments proposed in this paper are illustrated on small graphs based on fisheries data. Joo et al[65] identified pairwise relationships between vessels of some fishing fleets (groups of vessels sharing the same technical characteristics) based on joint-movement analysis[105] of their GPS-tracks at sea. On the basis of this previous work, we derive a set of twenty graphs describing the relationships (the edges) between a set of trips (the nodes) within a fleet of French trawlers with unknown topological properties and of unknown types.

Results and Discussion

Same orders and densities

The domain of separability between different models amongst $\{ER, SF, SW, GO\}$ is globally parabolic with regards to the order and the density. The range of edge densities allowing clear discrimination increases with the graphs' order. For instance, when comparing graphs coming from Erdős-Rényi (ER) and Fitness Scale-Free (SF) models (Fig 3.13a), for orders of 25 and 50, the domain of separability respectively spans a range of edge densities approximately from 0.3 to 0.6 and from 0.1 to 0.95, respectively. Furthermore, a perfect discrimination ($AUPR = 1$) is gradually reached for graphs with more than 50 nodes, more and more irrespective of the edge density.

The parabolic shapes of the surfaces of the domain of separability described by the AUPR isolines are qualitatively similar but not identical regardless of which pair of models are compared (S1 Figa-e), with the exception of the Erdős-Rényi (ER) vs. the Geometric (GO) model (S1 Figa), where the domain of separability exhibits a strong asymmetrical surface with regards to the density. The combination of all domains of separability (Fig 3.13b) appears as a mixture of all the domains of separability of each pair of models. The asymmetrical domain of separability between Erdős-Rényi and Geometric model (S1 Figa) is easily recognisable between the densities from 0.1 to 0.2, as well as the one between Erdős-Rényi and Watts-Strogatz small-world model (S1 Figb) for densities from 0.7 to 0.9.

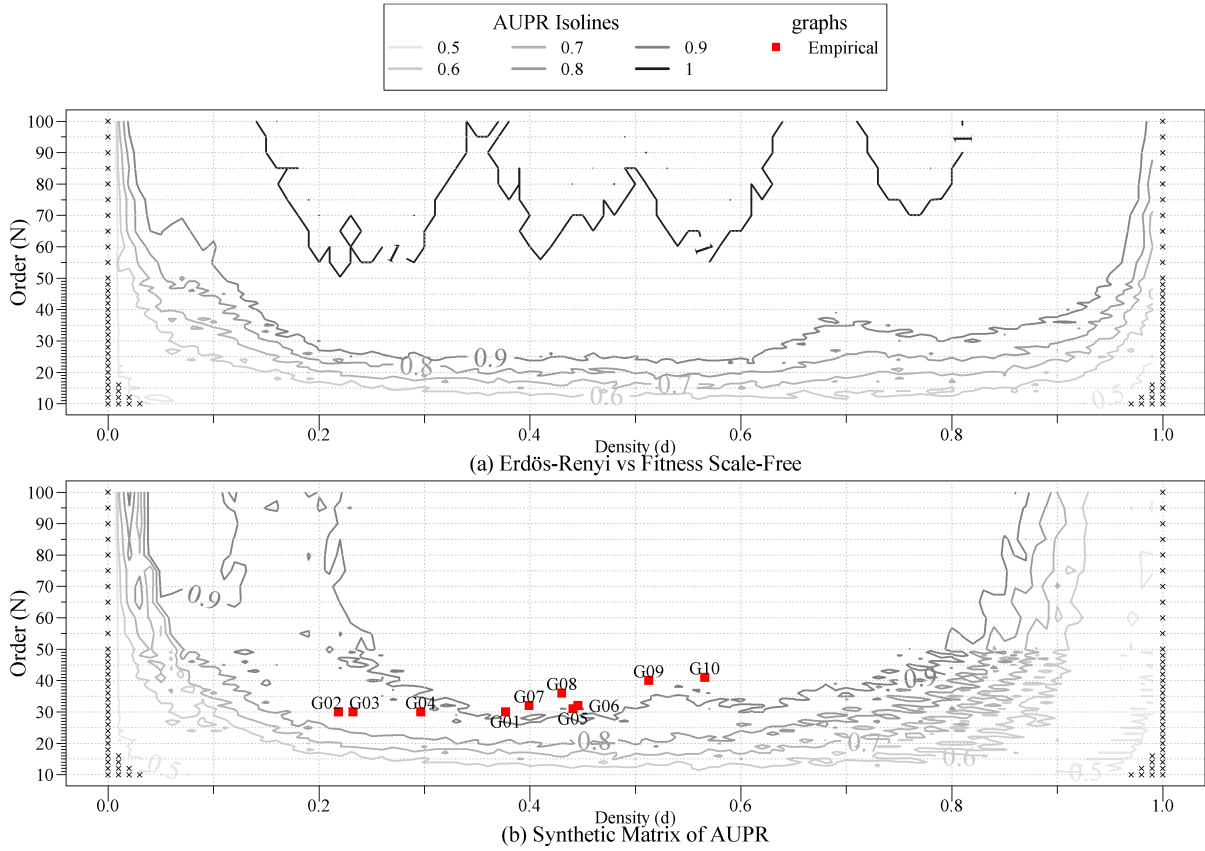


FIGURE 3.13 – **Quality of the discrimination between models (AUPR).** (a) Diagram of the matrix of AUPR Erdős-Rényi vs Fitness Scale-Free ($A_{ER,SF}$), (b) Diagram of the synthetic matrix of AUPR ($A(N, d)$). For each pair of models, and for each order (from 10 to 100) and edge density (from 0 to 1) combination, the quality of clustering between 100 graphs of each of the two types of models is assessed by the Area Under the Precision-Recall curve (AUPR). A maximum value of 1 corresponds to perfect discrimination. Black crosses represent zero AUPR. Empirical graphs (red squares) are projected according to their features (order and edge density).

The effect of the order and the density on the performance of the GCD_{11} are related to the response of the different graphlet correlation coefficients to changes in order and density (Fig 3.14). For a given density, the variability of each graphlet correlation coefficient is very high for small orders (lower triangle on Fig 3.14) leading to a small difference between the graphlet correlation coefficients of the four different models which are strongly overlapping. With increasing order (upper triangle on Fig 3.14), the variability of the graphlet correlation coefficients decreases leading to better discrimination (the domain of separability) between the four different models notably for some pairs of orbits in ramified graphlets of order 4 from O_6 to O_{11} , (except $O_7 \times O_{11}$). In other words,

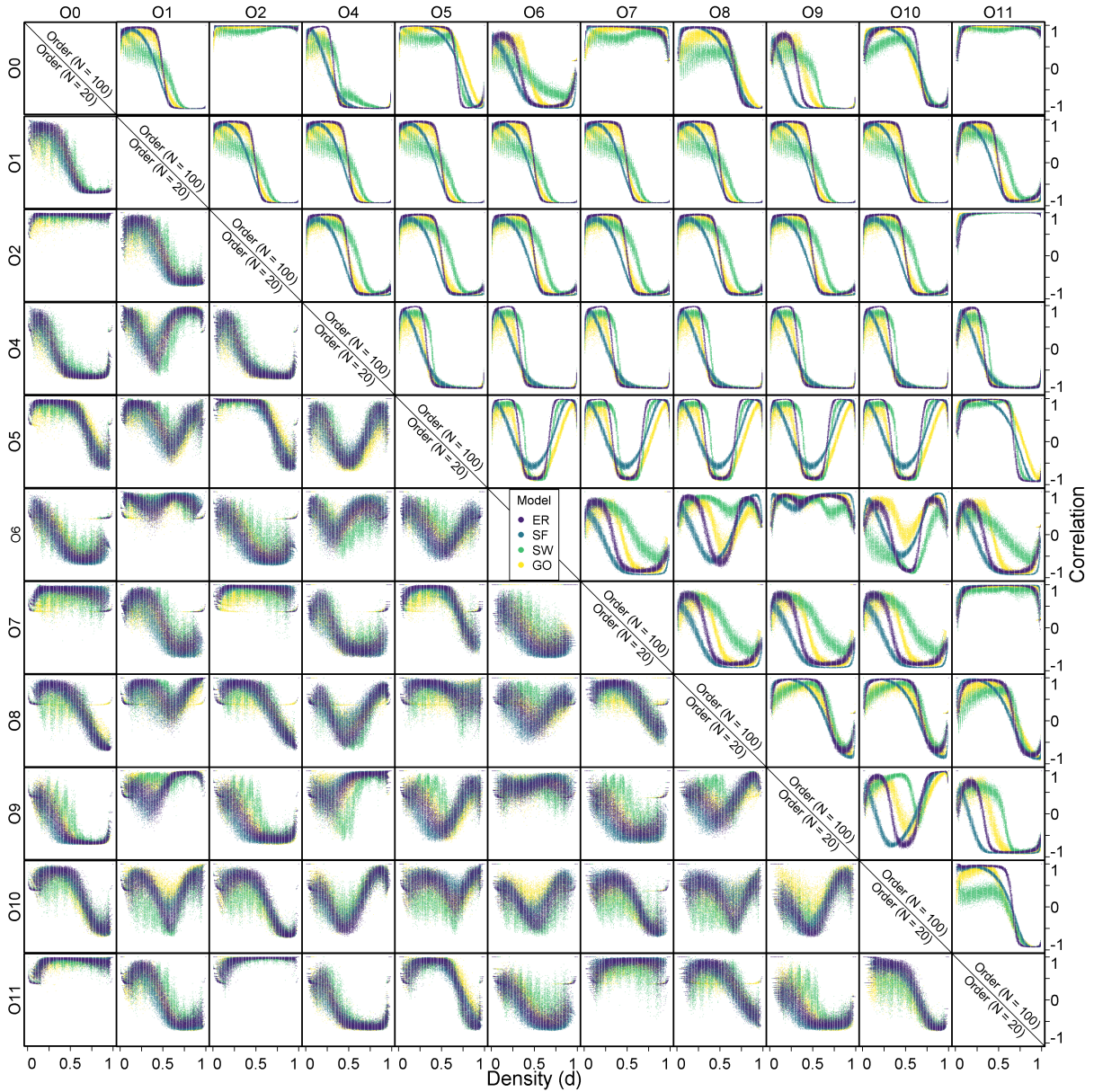


FIGURE 3.14 – **Evolution of the 55 graphlet correlation coefficients.** Erdős-Rényi (*ER*, in dark purple), Fitness Scale-Free (*SF*, in blue), Watts-Strogatz small-world (*SW*, in green) and Geometric (*GO*, in yellow) models. For two different order values $N = 20$ (lower triangle figures) and $N = 100$ (upper triangle figures), the graphlet correlation coefficients are computed for 100 graphs of the four models and for edges densities ranging from 0 to 1.

the increase in order allows to stabilise the graphlet correlation coefficients. Two reasons could explain this phenomenon. On one hand, the increase in order allows the emergence of complex topologies signing the topological properties of the model [156]. And finally,

Spearman's Correlation coefficient becomes more accurate when computed on a larger number of nodes [157].

For a given pair of orbits, the "starting point" from an empty graph (density $d = 0$) and the "ending point" at a complete graph (density $d = 1$) are at the same values of correlation coefficient regardless the graph model (isomorphic graphs [158], black crosses Fig 3.13). The graphlet correlation coefficients are thus strongly similar amongst models when the density approaches those two extremes.

Different orders and densities

When dealing with different orders and densities, the domain of separability of the GCD_{11} turns out to depend first on the order. For instance, when comparing graphs coming from Erdős-Rényi (ER) and Fitness Scale-Free (SF) models (Fig 3.15a), for equal orders (Fig 3.15b, block diagrams on the first bisector), the surface of the domain of separability increases from 0.045 to 0.19 when the order increases from 30 to 100. As with the same orders and densities, this means that due to reduced variability of the graphlet correlation coefficients, the edge density difference allowing clear discrimination between ER and SF is larger for "large" graphs. However, even for graphs with the same order, the difference in edge density allowing clear discrimination remains limited (Fig 3.15a).

This can be further detailed by principal component analyses (PCA) computed for a few cases taken as examples. Erdős-Rényi and Fitness Scale-Free graphs of order $N = 100$ and density $d = 0.2$ are compared with similar graphs of order $N = 100$ and densities in $d' = \{0.4, 0.6, 0.8\}$. In each case, we consider the matrix 400×55 whose lines are the 55 graphlet correlation coefficients (upper or lower triangle of the GCM) for each of the 100 simulated graphs of each type and of each density. Over the three cases considered, the first two principal components explain at least 96% of the input variability (Fig3.16). There are two orthogonal linear combinations of the graphlet correlation coefficients that represent the same amount of information as the 55 coefficients. Given that the 2D space of the two principal components contains nearly 100% of the initial variability, the euclidean distance in this 2D space is a suitable proxy for the GCD_{11} . Even a small difference in density (Fig3.16b) leads to the division of the graphs into four groups, discriminating first the model and secondly the density. As the density discrepancy increases (Fig3.16c-d), the denser graphs of the two different models are closer than the less dense ones.

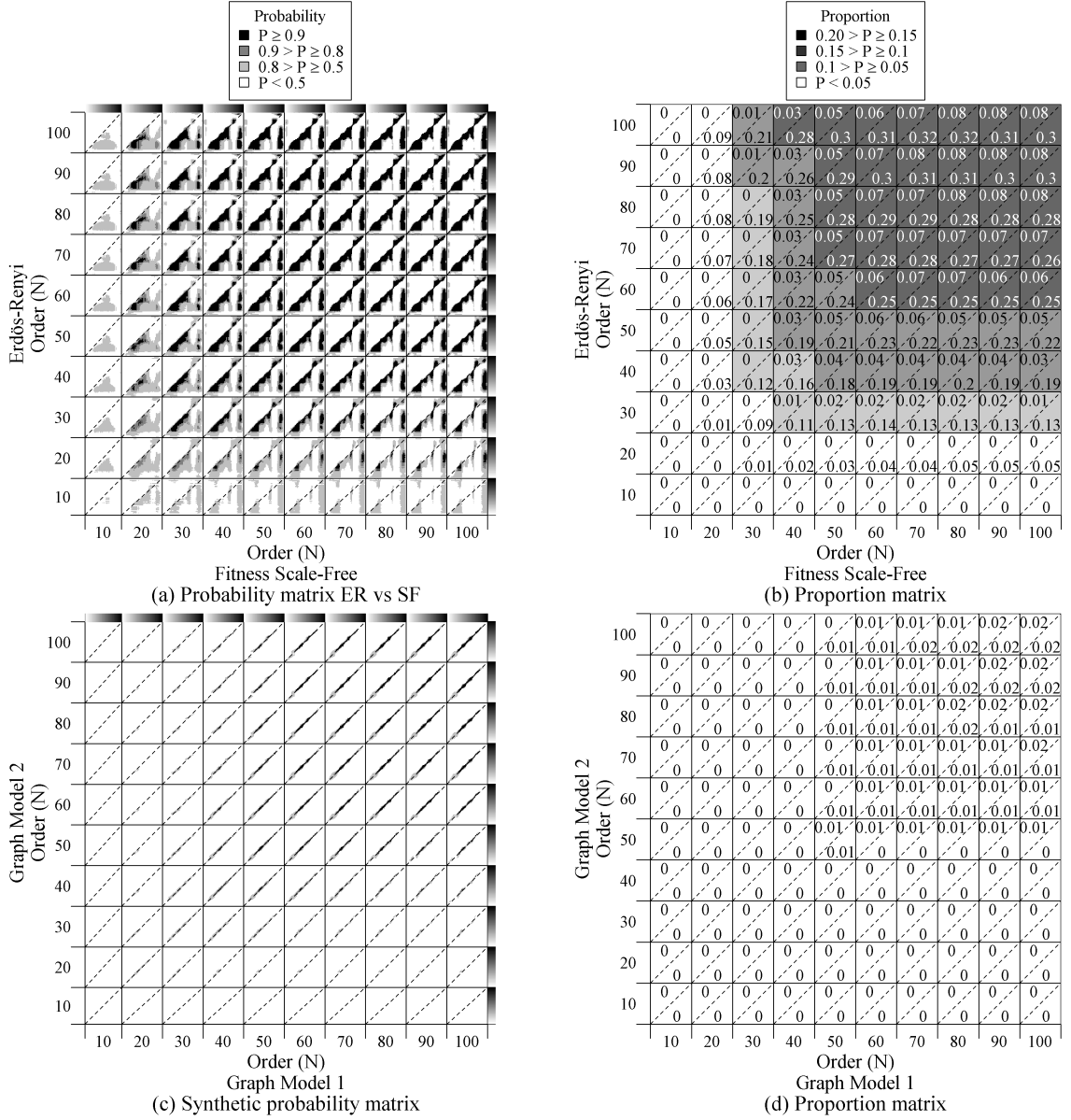


FIGURE 3.15 – **Probability of correctly distinguishing two random graph models with different order and/or edge density.** Each block (i, j) concerns the comparison of an M_1 of order N and a M_2 of order N' , with edge density d and d' respectively ranging from 0 to 1 (grey gradient from white to black on the top and right side) with $(M_1, M_2) \in \{ER, SF, SW, GO\}^2$ and $M_1 \neq M_2$. Dashed lines in each block highlight comparison when $d = d'$. (a) Probability matrix B_{M_1, M_2} that $D_{M_1(N, d), M_2(N', d')} > \max(D_{M_1(N, d), M_1(N', d')}, D_{M_2(N, d), M_2(N', d')})$ with $M_1 = SF$ and $M_2 = ER$. (b) Proportion of cells with a probability $P \geq 0.9$ in the triangle under or above the diagonal (cells covered by diagonals are not counted). Their mean quantifies the surface of the domain of separability of the GCD_{11} . (c) Synthetic matrix of probability matrix $B((N, d), (N', d'))$ according to each pair tested $(M_1, M_2) \in \{ER, SF, SW, GO\}^2$. (d) Proportion of cells in the synthetic matrix of probability with a probability $P \geq 0.9$.

This increasing similarity is explained by the growing similarity between the graphlets correlation coefficients with the density (Fig 3.14). The distances between the graphs associated with two different densities for a given model are larger than those between the two different models. In other words, the GCD_{11} does not discriminate graphs from different models with different densities because their topological differences do not depart more than the topologies of the same graph for different densities.

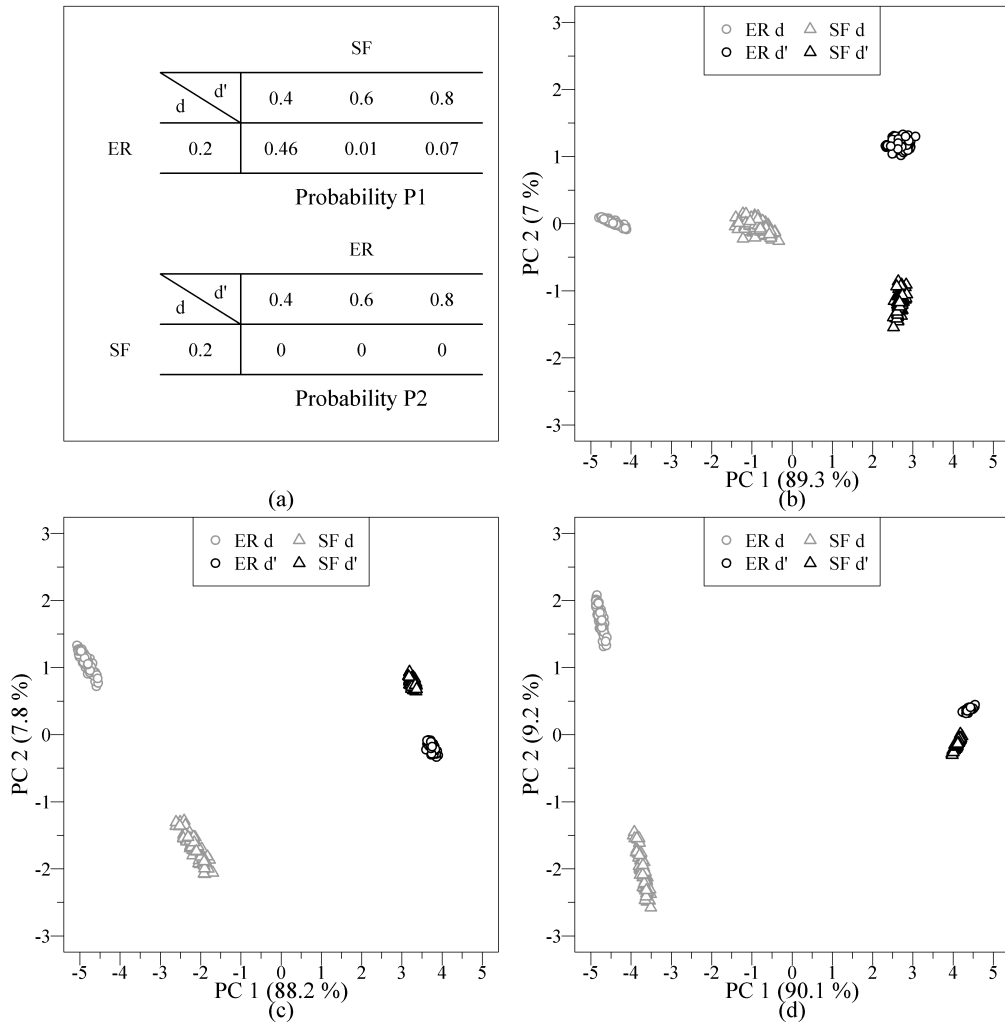


FIGURE 3.16 – PCA between Erdős-Rényi and Fitness Scale-Free graphs of order $N = 100$ and different edge density. (b) ($d = 0.2, d' = 0.4$), (c) ($d = 0.2, d' = 0.6$), (d) ($d = 0.2, d' = 0.8$). For each density, 100 graphs of each model are projected according to their 55 graphlet correlation coefficients (upper or lower triangle of the GCM). (a) The two probabilities P associated with each comparison in the matrix of probability $B_{ER,SF}$ are respectively $P1$ for $B_{ER(100,d),SF(100,d')}$ and $P2$ for $B_{ER(100,d'),SF(100,d)}$. With increasing edge density differences (b),(c) and (d) the two groups of dense graphs gradually become more and more similar.

Let us consider the reference comparison cases where the two graphs are of the same order (block diagrams on the first bisector in Fig 3.15b). Increasing the order for one of the two models by ten leads systematically to larger domains of separability and this increase is amplified when increasing the order of the *ER* model. For instance, starting with the comparison between *ER*(30, .) and *SF*(30, .) the domain of separability equal to 0.045, then, the domain of separability expands from 0.045 to 0.11 when the order of the *ER* graph increases (in columns), while it flattens around 0.07 when the increase of order concerns the *SF* graph (in rows).

The domain of separability is also systematically asymmetric favouring situations where the edge density of the *SF* graph is larger than the edge density of the *ER* graph it is compared to, whatever their respective orders (Fig 3.15a). The exhibited asymmetry of the domain is, however, dependent on the edge densities. As a matter of fact, when the orders increase, the domain of separability acquires a "violin" shape consisting of a body for the lower half range of edge density, a head for high or very high density, and a neck that appears as a transition between the body and the head.

While the comparison between the *ER* and *SF* model produces a domain of separability with a violin shape, the other model comparisons produce domains of separability with other shapes (S2 Fig) and surface. The largest separability domain is obtained when comparing Fitness Scale-Free and Watt-Strogatz small-world models (S2 Fige) with a strong asymmetrical shape favouring a situation where the edge density of the *SF* graph is larger than that of the *SW*. These results differ from those of the same order and density (Fig 3.13 and S1 Fig) where the diagrams of different pairs of models were quite similar. Due to this strong variability, the synthetic domain of separability $B((N, d), (N', d'))$ (Fig 3.15c) obtained by crossing all of the two-by-two probability matrices B_{M_1, M_2} is symmetric, very small and concentrated around the same density comparisons. Based on these results, the use of the GCD_{11} seems relevant only when the densities are very close. Furthermore, at least an order of $N = 50$ is required to obtain a domain of separability.

Empirical graphs comparison

Empirical graphs features

The empirical graphs used in this study are characterised by small orders ranging from 30 to 41 nodes and large edge densities ranging from 0.22 to 0.57 (Table.3.3). These empirical graphs (Fig 3.17c) show multiple dense components as a graph of communities.

Despite the short observation time (1 week), these graphs are sensitive to finer-scale temporal variations with some trips occurring at the beginning of the week and others at the end prohibiting an encounter between them.

TABLE 3.3 – **Main features of empirical graphs** : order (number of nodes), size (number of edges), and edge density (ratio between the size and the graph maximum size).

Graph	Order (N)	Size (S)	Density (d)
Graph 01	30	164	0.38
Graph 02	30	101	0.23
Graph 03	30	95	0.22
Graph 04	30	129	0.3
Graph 05	31	205	0.44
Graph 06	32	221	0.45
Graph 07	32	198	0.4
Graph 08	36	271	0.43
Graph 09	40	400	0.51
Graph 10	41	464	0.57
Mean	33.2	224.8	0.39
Range	[30 ; 41]	[95 ; 464]	[0.22 ; 0.57]

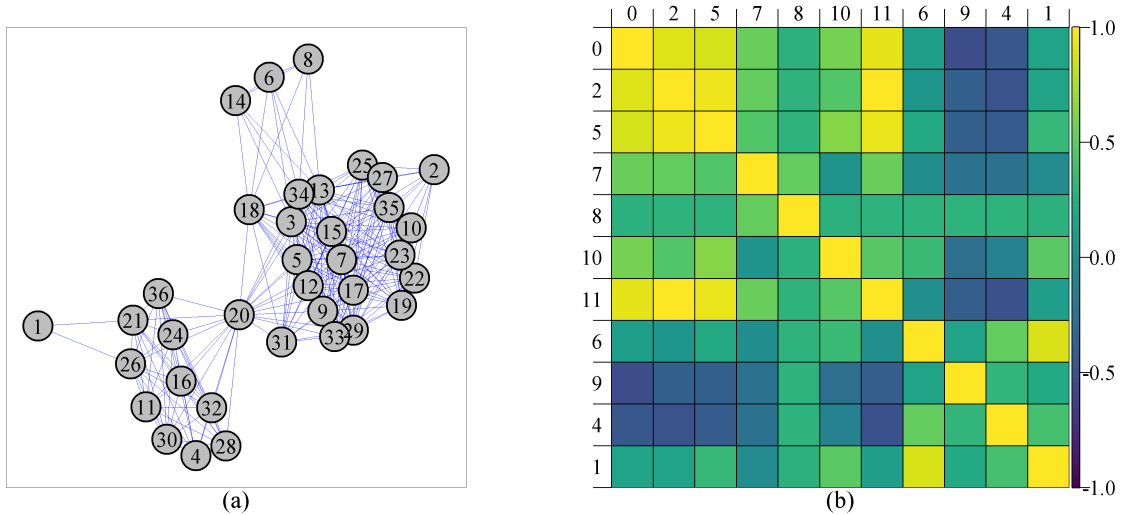


FIGURE 3.17 – **Illustration of an empirical graph and its Graphlet Correlation Matrix**. (a) Example of an empirical graph (Graph 08) and (b) its Graphlet Correlation Matrix (*GCM*). The 11 non-redundant orbits are grouped according to their *role*, orbit $\{0\}$ represents the familiar degree, $\{2, 5, 7\}$ represent node in chain, $\{8, 10, 11\}$ represent node in cycle, and $\{6, 9, 4, 1\}$ represent terminal node. Cell colours correspond to the value of the correlation coefficient between the 11 non-redundant orbits from 1 (yellow) to -1 (dark purple).

The *GCM* of empirical graphs (Fig 3.17d) exhibits a standard shape[132] with strong positive and negative correlations between the first eleven non-redundant orbits. These

contrasted correlations capture heterogeneity in the role of trips (nodes) in the graph. For instance, the negative correlations between orbits $\{4, 6, 9\}$ and orbits $\{0, 2, 5, 7, 8, 10, 11\}$ indicates the existence of peripheral nodes [132] such as nodes $\{1, 6, 8, 14\}$ which could reflect the solitary behaviour of some vessels. Conversely, the strong positive correlations between orbits $\{0, 2, 5, 7\}$ reflect the existence of a hub such as node 20 [132] which could reflect a strong sociability behaviour.

Testing model type

Due to their small orders, some of the empirical graphs (Graph $\{02; 03; 04\}$, red squares in Fig 3.13b) are out of the synthetic domain of separability. However, being outside the domain of separability does not preclude the use of the test. Being outside of the domain of separability means that it is difficult to systematically separate graphs coming from different models. However, the rejection of the null hypothesis, which reflects the strong dissimilarity between the empirical graph and the null model, remains valid.

According to our statistical test, none of the empirical graphs present any similarity with the same order and density graphs coming from Erdős-Rényi, Fitness Scale-Free and Watt-Strogatz small-world models (Table 3.4).

TABLE 3.4 – **Estimated p-values.** Each empirical graph is associated with an estimated p – value (\hat{p}) of being an outcome of an Erdős-Rényi, Fitness scale-free model, a Watts-Strogatz small word or a Geometric model. As in Table 3.3, empirical graphs are sorted according to their order. ($\hat{p}^* < 0.05$, $\hat{p}^{**} < 0.01$ and $\hat{p}^{***} \leq 0.001$)

Graph	p-value			
	Erdős Renyi	Scale Free	Watts-Strogatz	Geometric
Graph 01	0.001***	0.001***	0.001***	0.051
Graph 02	0.002**	0.001***	0.002**	0.122
Graph 03	0.001***	0.001***	0.001***	0.088
Graph 04	0.008**	0.032*	0.001***	0.029*
Graph 05	0.001***	0.001***	0.001***	0.154
Graph 06	0.001***	0.001***	0.001***	0.002**
Graph 07	0.001***	0.001***	0.001***	0.002**
Graph 08	0.001***	0.001***	0.001***	0.009**
Graph 09	0.001***	0.001***	0.001***	0.011*
Graph 10	0.001***	0.001***	0.001***	0.337

In terms of graphlet correlation coefficients, while there are similarities between the empirical *GCM* and the *GCMs* of the null model, for instance, Erdős-Rényi (Fig 3.18b), some empirical graphlet correlations are very different from those of the Erdős-Rényi. Interestingly these strong differences occur in orbits of high-order graphlets and specifically in

the pair of orbits which contain the orbit O_{10} , for instance, (O_9, O_{10}) , (O_4, O_{10}) , (O_1, O_{10}) , (O_2, O_{10}) , or (O_{10}, O_{11}) . A deep investigation of the role of orbit O_{10} could help to understand the topological differences of these graphs.

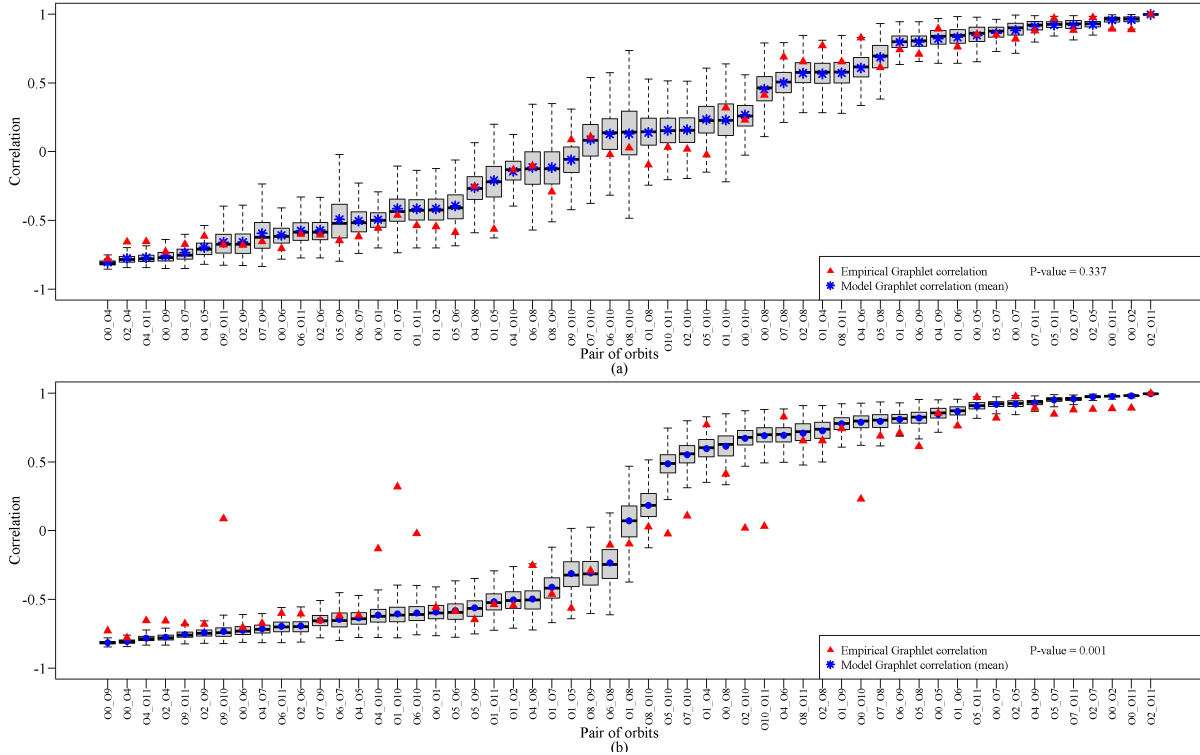


FIGURE 3.18 – Comparison between graphlet correlation coefficient of empirical and random graph model. (a) Graph 10 and Geometric model (b) Graph 10 and Erdős-Rényi model. For each pair of orbits, the graphlet correlation coefficients of 1000 Erdős-Rényi or Geometric graphs are presented as a boxplot with the mean value (blue asterisk) and the empirical value (red triangle). The correlation coefficients are mostly different from Erdős-Rényi graphlet correlation coefficients. Conversely, there is a strong similarity between graphlet correlation coefficients from Geometric graphs.

The comparison between the empirical graphs and graphs coming from the Geometric model leads to more contrasting results. While empirical graphs $\{04, 06, 07, 08, 09\}$ are significantly not similar to Geometric graphs, the test does not reject that empirical graphs $\{01, 02, 03, 05, 10\}$ can be considered as outcomes of a Geometric graph mode. However, empirical graphs 02 and 03 are out of the domain of separability which does not allow us to conclude that these graphs are geometric graphs. For empirical graphs 01 and 05 which are in the domain of separability, the small p -values (0.051 and 0.154), one being just above the threshold for rejection, can be interpreted in relation to the definition of

the domain of separability. Indeed, the AUPR larger than 0.9 associated with features of Graphs 01 and 05 (Fig 3.13b) implies a small, however not null, overlap between graphs coming from different models in $\{ER, SF, SW, GO\}$. This does not exclude the existence of “extreme” graphs from these models which might present some similarities. The largest p -value associated with Graph 10 suggests a strong similarity with Geometric graphs. Indeed, there is a strong similarity between most of the graphlet correlation coefficients of the empirical GCM of the graph 10 and the $GCMs$ of Geometric graphs (Fig 3.18a). This suggests that this empirical graph and Geometric graphs share similar topological properties.

To account for the contrasting variabilities of the correlation coefficients of each pair of orbits (boxplots in Fig 3.18), we also built a statistical test based on the standardised distance between $GCM(M_k)$ and \overline{GCM}_M (S1 Eq). This second test leads to similar conclusions (S1 Table) but rejects more clearly the null hypothesis. These results suggest that the standardised distance provides a more stringent test.

Conclusion

This work extends the use of the graphlet correlation distance (GCD_{11}) originally proposed for large real-world graphs to small real-world graphs. Through a numerical benchmark study based on four contrasted and commonly encountered random graph models, the Erdős-Rényi, Fitness scale-free, the Watts-Strogatz small world, and the Geometric graph models, we define the order \times density domain within which the Graphlet Correlation Distance (GCD_{11}) clearly separate graphs with the same order and the same density configuration. While the performance of the GCD_{11} is convincing to compare small graphs with the same order and the same density, when the orders and/or the densities differ, the performance of the GCD_{11} deteriorates quickly and exhibits a high variability depending on the pair of models compared. Therefore, it seems essential to systematically check the applicability of the GCD_{11} before comparing graphs of a different order and/or density to ensure the relevance and interpretability of its results.

For the four models used in this study, we show that the behaviours of the graphlets correlation coefficient in response to a change in density remain qualitatively similar. Furthermore, the proposed experimental design and numerical analysis can be directly used with other random graph models to explore new properties of the GCD_{11} and deeper investigate its domain of separability or to investigate other orbit combinations.

The statistical test and its standardised version proposed in this study to test the similarity between empirical graphs and graph models regardless of order and edge density can be applied without restriction on the size of the graphs. Some limitations of the GCD_{11} are highlighted on the basis of numerical evidence presented here.

The application to fisheries data originally intends to test whether certain fishing behaviours can be considered independent. This property is generally required to apply statistical inference methods and more particularly when estimating population biomass in marine ecosystems. However, our results suggest that the graphs issued from the GPS tracks are not similar to Erdős-Rényi graphs. Some of them are rather Geometric.

Finally, in the context of an application of the GCD on small graphs, the present work is also an opportunity to question the relevance of the choice of the 11 selected orbits for the construction of the GCM . Indeed, we have shown that there is a strong redundancy of the information provided by the different coefficients of graphlet correlations between pairs of orbits. Moreover, it seems that the information carried out by the correlation matrices (GCM) can be almost completely described in a two-dimensional space. We encourage further studies using a smaller number of not redundant orbits that could potentially improve the performance of GCD .

3.3.3 Application, l'indépendance des graphes de flottilles

Les contours du domaine de séparabilité de la GCD_{11} ont été esquissés, mettant en avant les gammes d'ordre et de densité des graphes permettant une comparaison dont les résultats peuvent être interprétés avec différents niveaux de confiance. De plus, cette sous-section a été le lieu du détournement de la GCD_{11} sous la forme d'un test statistique permettant de statuer sur la similarité d'un graphe (empirique ou non) au regard d'un ensemble de graphes (également empirique ou non) selon :

$$\begin{cases} H_0 : \text{Les graphes sont statistiquement similaires} \\ H_1 : \text{Les graphes ne sont pas statistiquement similaires} \end{cases}$$

Dans la sous-section précédente, seule une fraction des graphes empiriques de la flottille des OTB de Boulogne-sur-Mer a été présentée afin d'illustrer l'application de la GCD_{11} sous la forme du test statistique. Cette fraction ne permet pas de dresser une conclusion générale à l'échelle des deux années et des deux flottilles d'étude. De plus, les temps de publication étant ce qu'ils sont, les graphes empiriques utilisés dans cette publication sont légèrement différents de ceux considérés dans ce manuscrit du fait de la maturation de l'approche d'identification des dyades.

Dans cette sous-section, nous réitérons brièvement les analyses de la section précédente, appliquées à l'ensemble des graphes empiriques des deux flottilles, PTM de Cherbourg et OTB de Boulogne-sur-Mer.

L'indépendance des graphes de flottilles

Dans un premier temps, il s'agit d'identifier au regard des caractéristiques d'ordre N et de densité d de chaque graphe empirique, lesquels se situent dans le domaine de séparabilité de la GCD_{11} (Fig 3.19). Sans surprises, les 8 graphes de la flottille des PTM du fait de leur faibles ordres sont systématiquement en dehors du domaine de séparabilité. Concernant ceux de la flottille des OTB, malgré leurs ordres plus importants, seulement 8 graphes parmi les 94 ($\approx 8.5\%$) sont dans le domaine de séparabilité. Nous rappelons que les graphes empiriques sont ici des graphes où les noeuds représentent des trajectoires et les arêtes, une relation dyadique entre ces trajectoires (sans distinction de leur niveau de cohésion).

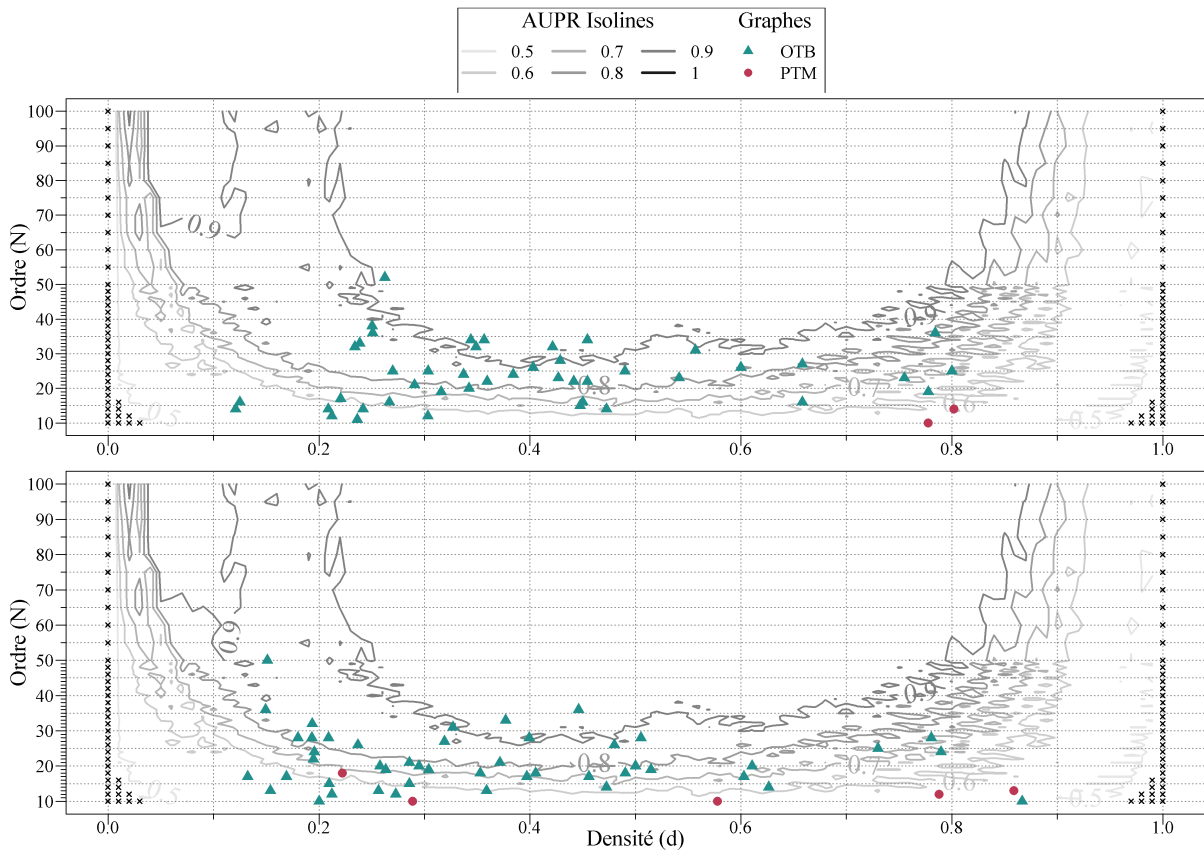


FIGURE 3.19 – Projection des 8 graphes empiriques de la flottille des PTM de Cherbourg et des 94 graphes empiriques de la flottille des OTB de Boulogne-sur-Mer dans le domaine de séparabilité de la GCD_{11} des années (a) 2012 et (b) 2013.

Afin de tester l'indépendance de l'organisation des relations dyadiques, chaque graphe empirique d'ordre N et de taille m est comparé au travers d'un test statistique à une collection de 1 000 réalisations d'un modèle d'Erdős-Rényi $G_{ER}(N, m)$. Au cours de ces comparaisons quelle que soit la position du graphe empirique dans le domaine de séparabilité de la GCD_{11} , le rejet de l'hypothèse nulle H_0 est toujours interprété comme une différence significative entre le graphe empirique et les réalisations du modèle d'Erdős-Rényi. A l'inverse, l'interprétation de la conservation de l'hypothèse nulle H_0 dépend de la position du graphe empirique dans le domaine de séparabilité de la GCD_{11} .

Ainsi, concernant les 8 graphes de la flottille PTM, le test rejette l'hypothèse nulle H_0 (p -value ≤ 0.05) (Fig 3.20). Concernant les graphes de la flottille OTB, il faut distinguer d'une part ceux relatifs à des graphes situés à l'intérieur du domaine de séparabilité, de ceux qui en sont en dehors. Du point de vu des premiers (8 graphes) le test rejette

l'hypothèse nulle H_0 . Du point de vu des seconds, le même résultat est observé pour environ les trois quarts des graphes (63 sur 86 soit $\approx 73\%$) tandis que pour les graphes constituant le quart restant, le test statistique conserve H_0 ($p - value > 0.05$).

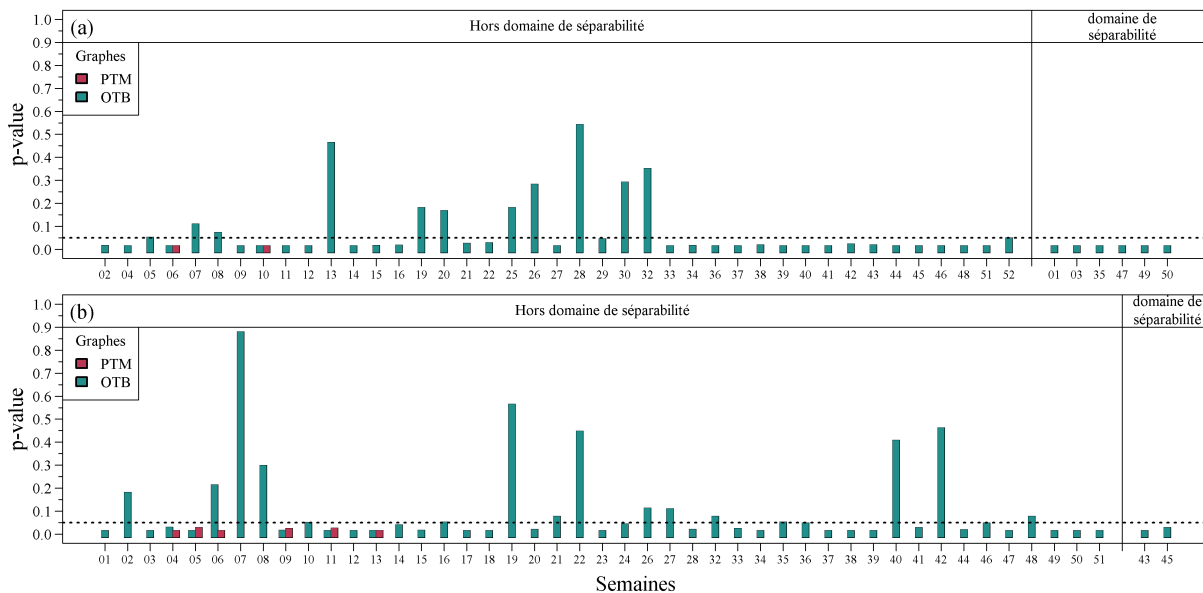


FIGURE 3.20 – P-value du test statistique de la comparaison entre les graphes empiriques OTB et les réalisations du modèle d’Erdős-Rényi de (a) 2012 et (b) 2013

Ainsi, nous montrons pour les deux flottilles PTM et OTB qu’il est statistiquement très improbable que l’organisation des relations dyadiques des graphes de trajectoires traduise des rencontres indépendantes .

3.4 Conclusion du chapitre

Au cours de ce chapitre, le choix d’aborder l’analyse de l’organisation topologique des relations dyadiques par des méthodes de comparaison de graphes nous a conduit à développer une approche méthodologique afin d’une part, d’interroger les performances de la GCD_{11} dans le monde des petits graphes, et d’autre part, de développer un test statistique.

Au terme de ce chapitre, même s’il s’agit d’un échantillon limité, nous montrons que l’organisation des relations dyadiques s’établissant entre les trajectoires des navires de pêche ne peut pas être jugée similaire à celle attendue sous l’hypothèse d’indépendance. Toutefois, la non-indépendance de l’organisation des relations dyadiques est-elle suffisante

pour affirmer de la non indépendance des trajectoires qui en sont à l'origine? C'est en partie à cette question que doit répondre le chapitre suivant.

DE L'INDÉPENDANCE DES RELATIONS À L'INDÉPENDANCE DES TRAJECTOIRES

Au cours du chapitre précédent, il a été question de développer une méthode permettant d'interroger l'organisation topologique des relations dyadiques s'établissant entre un ensemble de trajectoires de navires de pêche. Au travers de cette méthode reposant sur la comparaison entre des graphes empiriques et des réalisations de modèles de graphes aléatoires, il a été montré que l'organisation des relations dyadiques n'est pas représentative de celle attendue sous l'hypothèse d'indépendance. Toutefois, ces résultats sont insuffisants pour affirmer la non-indépendance des trajectoires sous-jacentes à ces relations dyadiques elles, non-indépendantes. En effet, le développement du chapitre précédent est articulé autour d'une hypothèse faisant le lien entre l'indépendance des trajectoires et l'indépendance des relations deux à deux. Dans ce chapitre, nous nous intéressons à la validité de cette hypothèse de travail.

4.1 Système d'hypothèse et démarche :

L'enjeu de ce dernier chapitre réside dans la validation ou la réfutation de l'hypothèse de travail faisant le lien entre l'indépendance des relations dyadiques et l'indépendance des trajectoires. Dans le premier cas, sa validation permettrait de conclure à la non-indépendance des trajectoires empiriques, tandis que dans le second, sa réfutation laisserait vacante toute conclusion, ouvrant la voie vers le développement d'un nouveau cadre d'analyse sous-tendu par de nouvelles hypothèses de travail.

La représentation d'un ensemble de trajectoires empiriques sous la forme d'un graphe de ses relations dyadiques a ouvert la voie aux hypothèses liées à l'organisation topologique de ses arêtes et à la comparaison aux modèles de graphes aléatoires. Le recours à ces hypothèses et à ces modèles de graphes est une réponse détournée à l'absence d'une

illustration de référence de ce que doit être un ensemble de trajectoires indépendantes. En effet, des ensembles de trajectoires dont l'indépendance est vérifiée pourraient être traduits en graphes de leur relations dyadiques, offrant une illustration de référence de graphes des relations 2 à 2 entre trajectoires indépendantes. Ainsi, en s'affranchissant de toute hypothèse à priori sur leurs propriétés, l'organisation topologique de ces graphes serait représentative de celle attendue sous l'hypothèse d'indépendances des trajectoires.

En disposant de tels graphes, il serait possible d'une part, d'interroger l'indépendance de leurs arêtes permettant ainsi de valider ou de réfuter l'hypothèse de travail reliant l'indépendance des trajectoires à celle des relations dyadiques, et d'autre part, en substituant les réalisations des modèles de graphes aléatoires par ces graphes de référence comme modèle nul, d'interroger la similarité de leur organisation topologique à celle des graphes empiriques de flottilles. Sur la base de ces deux considérations, un dernier système d'hypothèse est construit, cette fois-ci en deux parties :

Systeme d'hypothese n°3

Hypothèse : L'indépendance des trajectoires n'induit pas l'indépendance des relations dyadiques.

Prédiction n°1 : L'organisation topologique des relations dyadiques de trajectoires indépendantes n'est pas similaire à celle attendue sous l'hypothèse d'indépendance des arêtes (Erdős-Renyi).

Expérimentation n°1 : Évaluation de la similarité entre des graphes de trajectoires indépendantes et des graphes d'indépendance (Erdős-Renyi).

Observation n°1 : Voir résultats section 4.3

Si Prédiction n°1 vraie :

Expérimentation n°2 : Évaluation de la similarité entre des graphes de trajectoires indépendantes et des graphes empiriques de flottilles.

Observation n°2 : Voir résultats section 4.4.

4.2 Vers des graphes de trajectoires indépendantes

Ce dernier chapitre s'ouvre en évoquant la possibilité d'utiliser des trajectoires dont l'indépendance est vérifiée, comme illustration de référence. En sachant que le coeur de

ce travail est justement le développement d'une approche devant permettre d'interroger l'indépendance des trajectoires, mentionner l'existence de trajectoires dont l'indépendance est vérifiée, doit susciter certaines interrogations, d'où proviennent ces trajectoires indépendantes ? Et comment leur indépendance peut elle être vérifiée ?

C'est dans l'origine de ces trajectoires indépendantes, que se situe l'astuce et l'originalité de ce dernier chapitre. Puisqu'il est en l'état impossible de vérifier l'indépendance d'un ensemble de trajectoires empiriques, il reste toutefois la possibilité d'essayer d'en simuler. Ainsi, après s'être intéressé au travers du chapitre précédent aux modèles de graphes, dans ce chapitre et plus précisément encore, dans cette section, il est question de se tourner vers les modèles de trajectoires, dont un en particulier, incarné par les modèles de réseaux antagonistes génératifs, les GAN ("*Generative Adversarial Network*" en anglais).

4.2.1 Le modèle des réseaux antagonistes génératifs (GAN)

La simulation de trajectoires, loin d'être une tâche anodine est cependant fréquemment rencontrée, notamment en écologie comportementale. Par exemple, dans le cadre d'approches théoriques, le recours à la simulation permet de fournir des jeux de données "contrôlés" permettant d'étudier les effets de la variation d'un ou de plusieurs paramètres. Ainsi, à la manière des modèles de graphes, les modèles de trajectoires permettent de produire des réalisations, qu'il s'agisse d'une trajectoire seule ou d'un ensemble de trajectoires, dont les propriétés peuvent être étudiées. Ou bien dans le cadre d'approches plus appliquées, comme celle qui sera développée ici, les modèles de trajectoires peuvent incarner le rôle de modèle nul dont les réalisations peuvent être comparées à celles observées à partir de données empiriques [159].

Afin de répondre à ces multiples applications, une myriade de modèles de trajectoires a vu le jour, reposant sur différentes approches allant des processus à base de marches aléatoires [160, 161] à l'utilisation de systèmes dynamiques de représentation d'état ("*state-space model*") [59], en passant par les systèmes d'équations différentielles [162] ou par les outils d'apprentissage automatique ("*Machine learning*"). Ici, seule l'approche de simulation par réseaux antagonistes génératifs (GAN) est naïvement et modestement présentée sans trop entrer dans le détail.

Les GAN représentent une catégorie de méthodes d'apprentissage automatique ("*Machine learning*") présentant la particularité de reposer sur des algorithmes d'apprentissage profond ("*deep-learning*") [163], caractérisés par une architecture en réseaux de neurones

artificiels. C'est en faisant passer au travers de ces réseaux de neurones des jeux de données (initialement des images ou des sons), que les méthodes d'apprentissage automatique parviennent à décomposer et à saisir les caractéristiques, parfois les plus complexes de ces jeux de données (Fig 4.1).

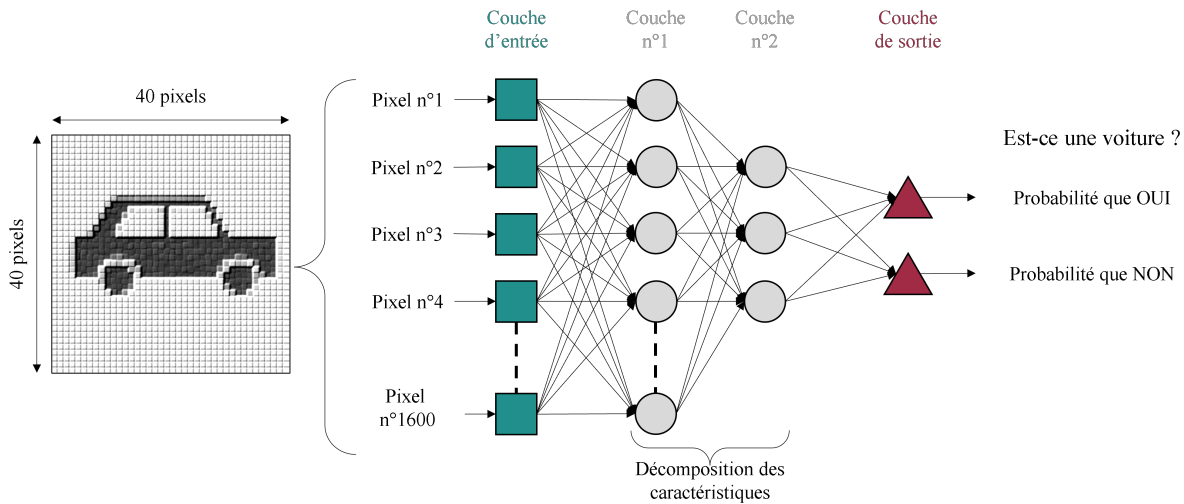


FIGURE 4.1 – Illustration de la méthode de classification par l'apprentissage profond.

Ces outils sont rapidement devenus des incontournables des applications de classification en représentant une alternative aux arbres de régression qui nécessitent au préalable la définition de variables discriminantes. De plus, l'architecture extrêmement modulable des réseaux de neurones vis-à-vis de leur nombre de couches et/ou du nombre de neurones présents et/ou de la nature des fonctions associées à chaque neurone, permet d'adapter les performances et la finesse de la description en fonction des données à traiter.

Les approches par GAN se distinguent des simples méthodes d'apprentissage pur par leur aspect génératif les rendant capables de simuler des données "semblables" à celles fournies à des fins d'apprentissage [164]. Pour assurer cette double capacité, d'apprentissage et de simulation, les GAN reposent sur une architecture composée de deux réseaux de neurones distincts, l'un appelé le générateur en charge de simuler des données, et l'autre appelé le discriminateur entraîné à reconnaître et à classer des données (Fig 4.2).

Dans cette configuration, ces deux réseaux sont organisés pour se livrer un affrontement sous la forme d'un jeu à somme nulle[165]. D'un côté, le réseau générateur a pour objectif de produire par simulation un "candidat", dans le cas présent une trajectoire, qu'il va soumettre au réseau discriminatoire. De l'autre, le réseau discriminatoire qui a été entraîné

à partir d'un jeu d'apprentissage à reconnaître les caractéristiques des données réelles (là aussi des trajectoires), doit évaluer si le candidat qui lui est présenté est selon lui réel ou simulé. En d'autres termes, l'objectif du réseau générateur est de réussir à proposer un candidat capable de tromper le réseau discriminatoire, permettant le cas échéant, de considérer ce candidat comme réaliste au regard des données réelles.

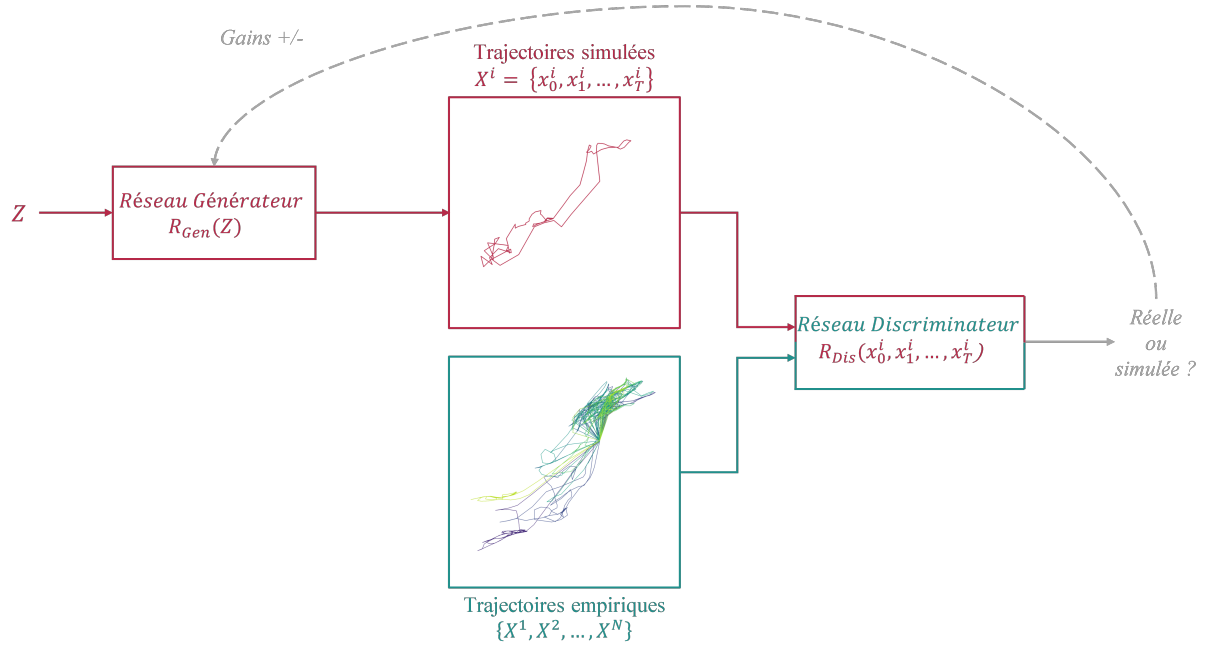


FIGURE 4.2 – Illustration de l'architecture d'un GAN et de son mécanisme de génération et de validation des données simulées (d'après [166]).

C'est la mise en place d'objectifs contradictoires entre ces deux réseaux qui est à la base des performances des approches par GAN, où le réseau générateur est poussé à apprendre en aveugle, à produire des candidats de plus en plus réalistes aux yeux du réseau discriminatoire. Là aussi, l'architecture des réseaux de neurones générateur et discriminatoire peut être adaptée selon les données à simuler. L'architecture du GAN utilisée dans ce travail est issue de la littérature et est empruntée d'un travail réalisé par Roy et al [166] en 2022, dans lequel, les auteurs ont eu recours au GAN afin de simuler des trajectoires d'oiseaux marins. Au delà du fait qu'il s'agisse d'un modèle de simulation de trajectoires, l'architecture du GAN développée par les auteurs a été spécifiquement élaborée au regard d'un type particulier de trajectoires de ces oiseaux marins, qualifiées de "trajectoires de recherche de nourritures à lieu central" ("*central-place foraging trajectories*"). S'agissant d'oiseaux nichant sur des îles, cette dénomination fait référence aux trajectoires réalisées

par les individus traduisant l'aller-retour de la colonie (le lieu central) vers les zones de nourriture, elles, situées en mer. Ainsi, de la même manière que les trajectoires des navires de pêche débutent et se terminent au même endroit (au même port), les trajectoires de ces oiseaux marins décrivent une boucle dont les points de départ et de retour sont confondus, incarnés par le lieu où niche la colonie. Sur cet aspect, les trajectoires des oiseaux et des navires de pêche apparaissent très similaires, renforçant la pertinence du choix de ce modèle.

4.2.2 Des trajectoires indépendantes au graphe

Le recours à l'utilisation d'un modèle de GAN doit permettre de simuler des trajectoires d'une part indépendantes, et d'autres part, réalistes au regard des trajectoires réelles d'apprentissage. Toutefois, comme toutes méthodes d'apprentissage automatique et plus largement comme toutes méthodes de classification, les performances attendues sont fortement conditionnées par la qualité du jeu d'apprentissage et en particulier par le nombre d'exemplaires fournis à l'algorithme. Pour prendre des exemples extrêmes, le système "*DeepFace*" de reconnaissance faciale développé par le groupe Facebook a été entraîné sur plus de quatre millions d'images [167], tandis que le programme "*AlphaGO*" appartenant aujourd'hui à Google, premier programme ayant réussi à vaincre un joueur professionnel au jeu de go, a quant à lui été entraîné sur 160 000 parties représentant au total, un ensemble de plus de 29 millions de coups joués [168].

Sélection des jeux d'apprentissage

Tout en conservant le regroupement des trajectoires empiriques sous la forme d'ensembles hebdomadaires, il s'agit de sélectionner parmi ces ensembles ceux présentant un nombre suffisant de trajectoires. Dans leurs travaux, Roy et al, disposaient de trois ensembles de trajectoires d'oiseaux contenant respectivement 30, 50 et 78 trajectoires [166]. Compte tenu de ces jeux de données, et au regard du plus grand ensemble de trajectoires empiriques de navires de pêche constitué de 45 trajectoires, il est arbitrairement fixé de ne conserver que les ensembles relatifs à des semaines contenant au minimum 30 trajectoires (Table 4.1). Par manque d'ensemble répondant favorablement à ces critères, la flottille des PTM est exclue de ce pan d'analyse, et nous avons exclusivement travaillé avec la flottille OTB de Boulogne-sur-Mer représentée par 19 ensembles de trajectoires, comprenant entre 30 et 45 trajectoires chacun.

TABLE 4.1 – Caractéristiques des ensembles de trajectoires sélectionnés de la flottille OTB.

Année	Semaine	Ensemble	Nombre de trajectoires	durée moyenne (h)	Nombre de dyades
2012	01	01	44	33	294
	03	02	31	72	250
	11	03	36	69	498
	14	04	30	70	251
	34	05	30	69	110
	35	06	34	67	174
	46	07	33	56	139
	47	08	31	53	177
	48	09	33	60	169
	49	10	31	60	207
	50	11	32	67	181
2013	05	12	32	47	152
	36	13	32	56	92
	37	14	45	50	177
	39	15	30	76	151
	43	16	31	59	234
	45	17	32	52	180
	46	18	32	54	91
	50	19	30	55	115

D'une trajectoire simulée à un ensemble de trajectoires

Chaque ensemble i de N^i trajectoires, noté $E_{emp}^i = \{X^1, X^2, \dots, X^{N^i}\}$ avec $i = 1, \dots, 19$ est associé à un modèle de GAN, noté GAN_i dont les réalisations (les trajectoires simulées) doivent rendre compte des caractéristiques des trajectoires propres à chaque ensemble E_{emp}^i . Pour chaque ensemble E_{emp}^i , le nombre de trajectoires simulées par le modèle de GAN associé GAN_i est fonction du nombre de trajectoires N^i selon $200 \times N^i$. Ainsi par exemple, pour l'ensemble n°08 et n°13 correspondant respectivement à la semaine 47 de l'année 2012 et à la semaine 36 de l'année 2013, c'est au total $31 \times 200 = 6200$ et $32 \times 200 = 6400$ trajectoires qui sont simulées. Considérée de manière isolée, il est difficile par une simple inspection visuelle de distinguer les trajectoires empiriques et les trajectoires simulées, traduisant le "réalisme" des simulations (Fig 4.3).

Il s'agit à présent de constituer des ensembles de trajectoires simulées aux caractéristiques collectives cohérentes vis-à-vis de celles obtenues empiriquement. Au delà du nombre de trajectoires N^i qu'il contient, un ensemble de trajectoires empiriques E_{emp}^i est également caractérisé d'une part, par la distribution des durées T de ses trajectoires et d'autres part, par leur fenêtre temporelle. Or comme cela a été évoqué au cours du

chapitre précédent (Section 3.3), la distribution temporelle des trajectoires est un facteur clefs de leur capacité à établir des relations dyadiques. Ainsi, afin de rendre la comparaison entre des ensembles de trajectoires empiriques et simulées la plus pertinente possible, il est important de s'assurer que ces deux types d'ensembles soient définis dans un espace temporel similaire.

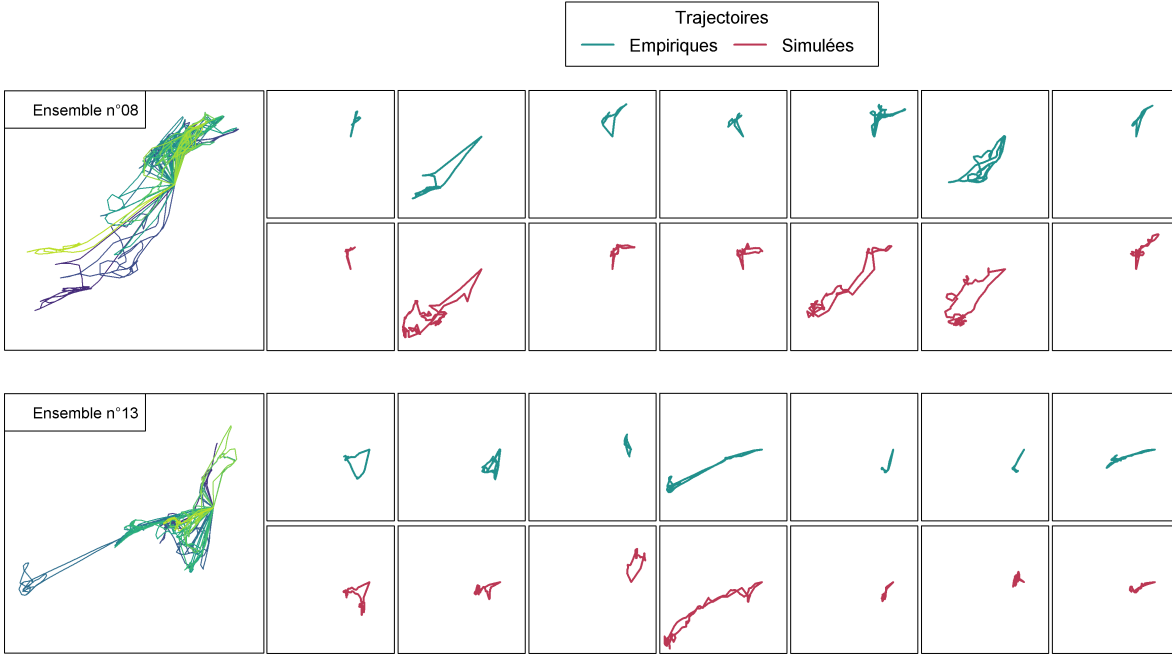


FIGURE 4.3 – Illustration de trajectoires empiriques de la flottille OTB et de trajectoires simulées par GAN.

Considérant un ensemble E_{emp}^i défini par son vecteur des durées de ses N^i trajectoires selon $T_{emp}^i = \{T_{emp}^{i,1} \leq T_{emp}^{i,2} \leq \dots T_{emp}^{i,N^i}\}$, il s'agit de constituer des ensembles de trajectoires simulées $E_{GAN}^{i,j}$ défini chacun par leur vecteur des périodes de leur N^i trajectoires $T_{GAN}^i = \{T_{GAN}^{i,1} \leq T_{GAN}^{i,2} \leq \dots, T_{GAN}^{i,N^i}\}$, tout en minimisant $|T_{emp}^{i,k} - T_{GAN}^{i,k}|$ avec $k = 1, \dots, N^i$.

Sur la base de ce critère, à partir d'un ensemble E_{emp}^i , il est arbitrairement convenu de constituer 100 ensembles de trajectoires simulées $E_{GAN}^{i,j}$ avec $j = 1, \dots, 100$. La simulation d'un nombre de trajectoires deux fois supérieur à celui nécessaire à la construction des 100 ensembles $E_{GAN}^{i,j}$ permet une marge de manoeuvre dans la sélection des trajectoires parmi la collection des $200 \times N^i$ trajectoires simulées afin de minimiser au mieux le critère temporel. En effet, les trajectoires simulées par un GAN peuvent présenter des durées variables dont certaines ne sont pas représentées au sein des trajectoires empiriques, par

exemple, une trajectoire simulée peut présenter une durée supérieure ou inférieure à la plus longue ou à la plus courte des trajectoires empiriques.

Au travers de cette répartition supervisée, il est possible de constituer des ensembles de trajectoires simulées $E_{GAN}^{i,j}$ présentant des caractéristiques globales extrêmement similaires à celles observées sur l'ensemble de trajectoires empiriques E_{emp}^i qui leur sont associées (Fig 4.4). Chaque ensemble j de $E_{GAN}^{i,j}$ est alors défini par un même nombre de trajectoires N^i , des durées similaires ou identiques (Fig 4.4a,d) et se chevauchant temporellement d'une manière similaire ou identique (Fig 4.4c,f).

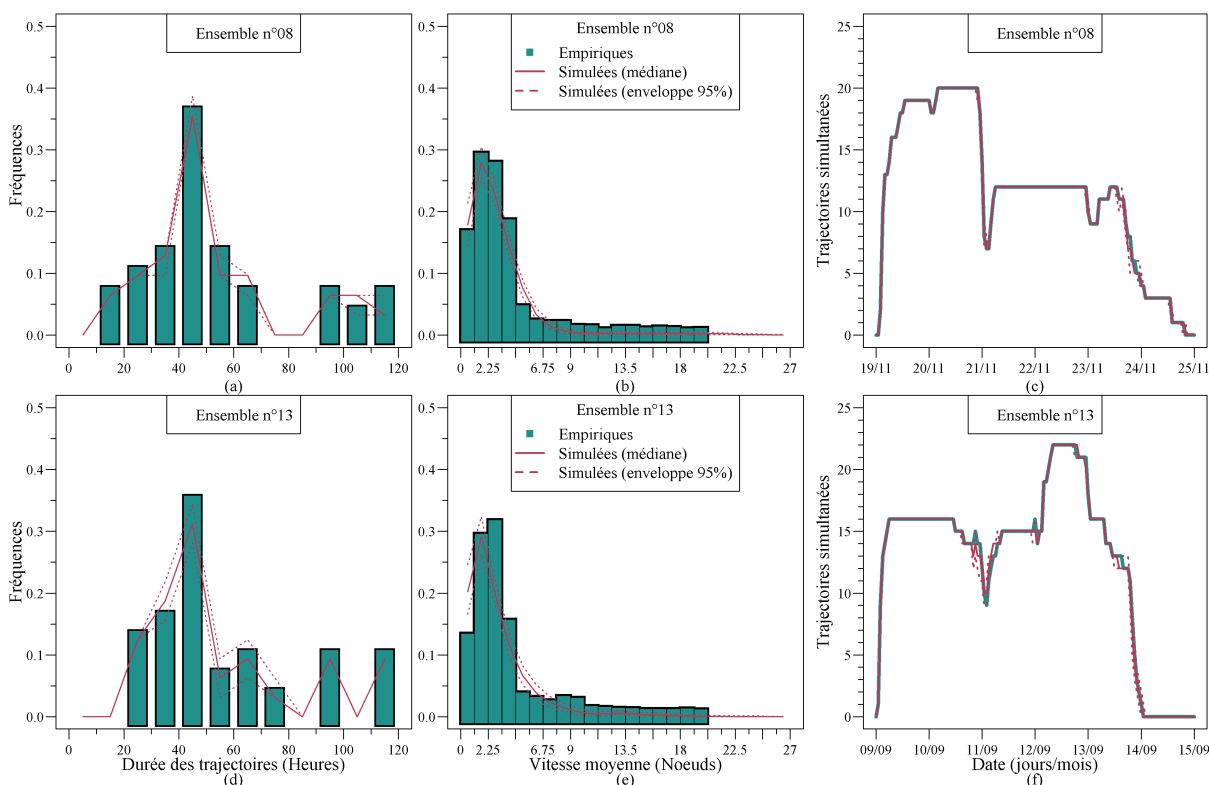


FIGURE 4.4 – Illustration des caractéristiques globales d'un ensemble de trajectoires empiriques et comparaison avec celles d'un ensemble de trajectoires simulées, (a) et (d) distributions des durées des marées, (b) et (e) distribution des vitesses moyennes entre deux pas de temps, (c) et (f) série temporelle du nombre de trajectoires simultanément en cours.

Indépendamment de ces considérations sur les caractéristiques globales des ensembles de trajectoires, liées à la manière dont sont réparties des trajectoires simulées, il est intéressant de noter que les ensembles de trajectoires simulées présentent également une distribution des vitesses moyennes proches de celle observée empiriquement (Fig 4.4b,e),

traduisant l'existence des deux régimes de vitesse, un lent ($\leq 4.5nds$) et un rapide ($> 4.5nds$), distinguant les déplacements lorsque l'engin de pêche est déployé ou non. La persistance de cette caractéristique dans les trajectoires simulées suggèrent que le GAN l'a identifiée comme une facette importante du réalisme des trajectoires.

En cherchant à reproduire le plus fidèlement possible les caractéristiques globales d'un ensemble de trajectoires empiriques, il est possible de supposer que les éventuelles différences qui pourraient être observées entre les graphes empiriques et les graphes de trajectoires indépendantes, soient uniquement le reflet d'une différence liée à l'organisation collective des trajectoires les unes par rapport aux autres. Avant d'interroger l'organisation des trajectoires simulées, au travers de leur graphes, il est possible d'inspecter visuellement cette organisation au regard de celle fournie par les trajectoires empiriques (Fig 4.5).

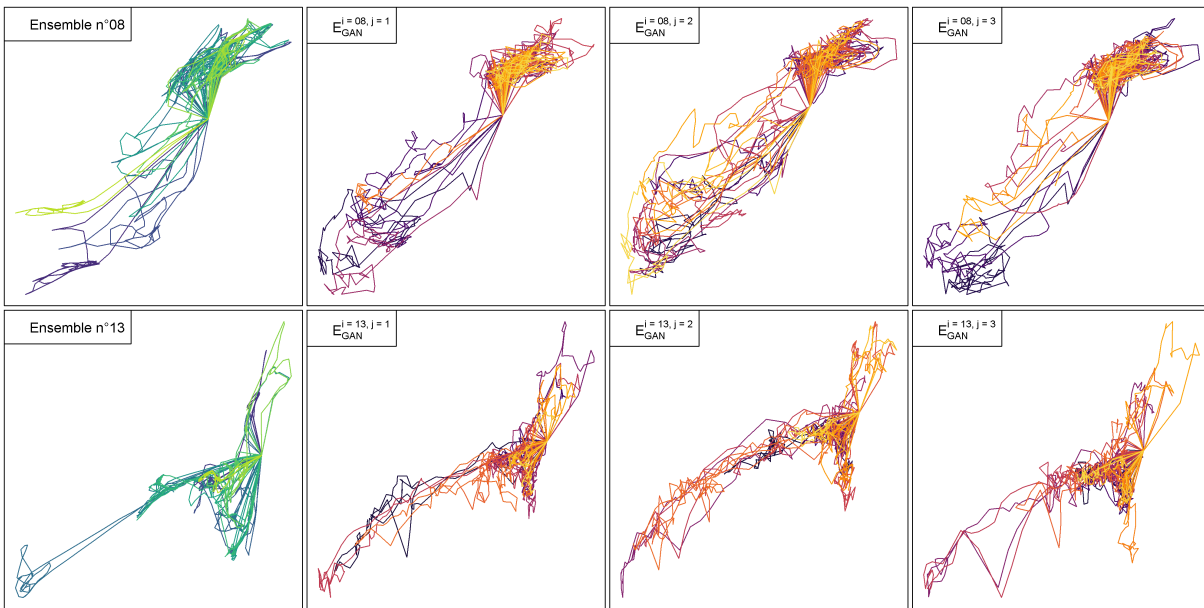


FIGURE 4.5 – Représentation en ligne de trois ensembles de trajectoires simulées (gradient jaune-violet) au regard du jeux de trajectoires empiriques d'apprentissage (gradient jaune-vert).

D'une manière générale, les ensembles de trajectoires indépendantes $E_{GAN}^{i,j}$ épousent plus ou moins finement les contours de la forme globale décrite par les trajectoires empiriques E_{emp}^i . Toutefois, sans véritablement être en mesure d'en définir la raison, les ensembles de trajectoires simulées apparaissent visuellement différents de ceux des trajectoires empiriques, ces dernières semblant s'assembler d'une manière moins erratique.

Autrement dit, les ensembles de trajectoires simulées d'un même modèle de GAN semblent plus similaires entre eux que lorsqu'ils sont comparés à l'ensemble des trajectoires empiriques qui en est à l'origine.

D'un ensemble de trajectoires simulées au graphes

Maintenant que des illustrations de ce que doivent être un ensemble de trajectoires indépendantes ont été produites, il s'agit de traduire ces illustrations sous la forme de graphes afin, une nouvelle fois, d'être en mesure d'interroger l'organisation des trajectoires les unes par rapport aux autres. Cette étape de traduction, est en tout point identique à celle effectuée sur les trajectoires empiriques.

Pour chaque ensemble j de trajectoires indépendantes $E_{GAN}^{i,j}$, chaque couple de trajectoires est considéré afin d'identifier l'ensemble des relations dyadiques, où il est rappelé ici, qu'une relation dyadique existe entre deux trajectoires, si elles se situent simultanément à une distance inférieure ou égale à 5 km l'une de l'autre. L'ensemble des relations dyadiques est enfin intégré sous la forme d'un graphe où chaque noeud représente une trajectoire et une arête entre deux noeuds, l'existence d'une relation dyadique (Fig 4.6).

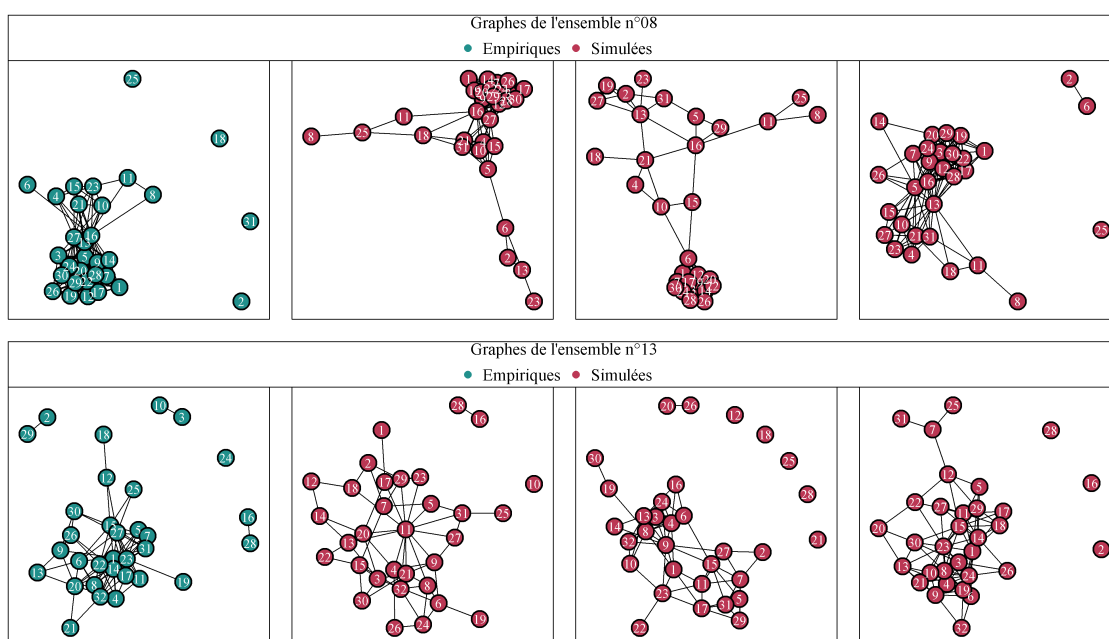


FIGURE 4.6 – Représentation en ligne de trois exemples de graphes de trajectoires indépendantes au regard des graphes de trajectoires empiriques.

Ainsi chaque ensemble de trajectoires empiriques E_{emp}^i décrit par un graphe de ses re-

lations dyadiques G_{emp}^i , est associé à une collection de 100 graphes décrivant l'organisation des relations dyadiques entre des trajectoires indépendantes $G_{GAN}^{i,j}$ avec $j = 1, \dots, 100$.

4.3 Graphes de trajectoires indépendantes *versus* graphes d'arêtes indépendantes (Erdős-Rényi)

Dans cette section, il s'agit de trancher en faveur de la validation ou de la réfutation de l'hypothèse liant l'indépendance des trajectoires à l'indépendance des relations dyadiques, hypothèse sous-tendant l'ensemble du chapitre précédent. De la même manière que cela a été réalisé concernant les graphes $G_{emp}(N, d)$ des trajectoires empiriques, le caractère indépendant des relations dyadiques s'établissant entre les trajectoires indépendantes (simulées) au travers d'un graphe $G_{GAN}(N, d)$ peut être interrogé au travers d'une comparaison avec des réalisations du modèle d'Erdős-Rényi $G_{ER}(N, d)$.

4.3.1 Graphes de trajectoires indépendantes et domaine de séparabilité

Contrairement au cas des trajectoires empiriques où chaque ensemble hebdomadaire E_{emp}^i est décrit par un unique graphe G_{emp}^i , il s'agit ici d'interroger les 100 graphes $G_{GAN}^{i,j}$ issus d'un même modèle G_{GAN}^i , $E_{GAN}^{i,j}$ avec $j = 1, \dots, 100$. Avant de pouvoir réaliser la comparaison entre chacun de ces 100 graphes et des réalisations d'un modèle d'Erdős-Rényi, il est nécessaire de s'assurer de la pertinence de cette comparaison au regard des caractéristiques d'ordre N et de densité d de chacun d'entre eux.

Pour chaque ensemble $E_{GAN}^{i,j}$, les 100 graphes $G_{GAN}^{i,j}$ sont ainsi projetés selon leur ordre $N^{i,j}$ et leur densité $d^{i,j}$ respectives, dans le domaine de séparabilité de la GCD_{11} (Fig 4.7). Au regard d'un même ensemble i , les 100 graphes $G_{GAN}^{i,j}$ peuvent présenter une certaine variabilité d'ordre N et de densité d . Concernant la variation de la densité, celle-ci est imputable aux aléas de l'organisation des trajectoires indépendantes les unes par rapport aux autres au sein de chaque ensemble j de $E_{GAN}^{i,j}$, influençant le nombre de relations dyadiques identifiées, et donc d'arêtes présentes dans chaque graphe $G_{GAN}^{i,j}$. Concernant la variation de l'ordre N , les 100 graphes $G_{GAN}^{i,j}$ d'un même ensemble i sont initialement construits avec le même nombre de noeuds N^i , relatif au nombre de trajectoires présentes empiriquement dans E_{emp}^i . Toutefois, comme cela a été fait pour les graphes empiriques, les

noeuds isolés traduisant des trajectoires n'intervenant pas dans des relations dyadiques, sont retirés des graphes pour la comparaison aux modèle d'Erdős-Rényi. Ici aussi, le nombre de noeuds isolés est variable selon les aléas de l'organisation des trajectoires indépendantes.

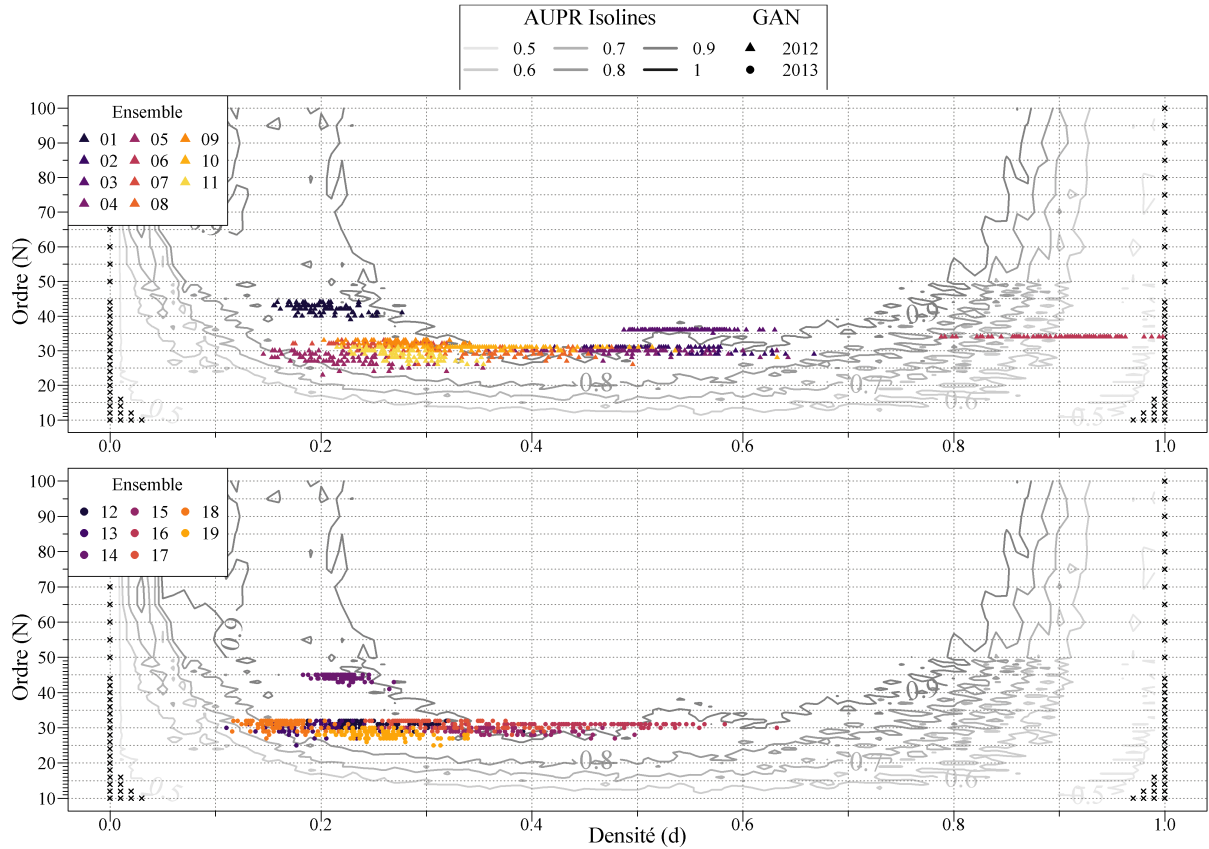


FIGURE 4.7 – Projection des graphes de trajectoires indépendantes dans le domaine de séparabilité de la GCD_{11} .

Du fait de cette variabilité de l'ordre et de la densité, la collection des 100 graphes $G_{GAN}^{i,j}$ issue d'un même modèle GAN_i , par exemple ceux de l'ensemble $i = 04$, peuvent être constitués à la fois de graphes situés à l'intérieur ($AUPR \geq 0.9$) et à l'extérieur ($AUPR < 0.9$) du domaine de séparabilité de la GCD_{11} . D'une manière générale, ces graphes se situent à proximité de la frontière du domaine de séparabilité ($0.8 < AUPR < 0.9$).

4.3.2 Comparaison aux graphes d'indépendance

Finalement, les 19×100 graphes de trajectoires indépendantes $G_{GAN}^{i,j}$ peuvent être comparés un à un à 1 000 réalisations du modèle d'Erdős-Rényi, de même ordre et de même densité $G_{ER}(N^{i,j}, d^{i,j})$. La similarité entre chaque graphe $G_{GAN}^{i,j}$ et les réalisations du modèle d'Erdős-Rényi est testée au travers du même test statistique que celui présenté au cours du chapitre précédent, caractérisé par une p -value statuant sur la similarité ou non de ces graphes avec p -value < 0.05 , H_0 est conservée, et p -value ≥ 0.05 H_0 est rejetée.

Pour chaque ensemble i , de 100 graphes j $G_{GAN}^{i,j}$, les 100 p -values sont séparées en deux groupes, distinguant celles obtenues sur des graphes situés à l'intérieur du domaine de séparabilité, des autres. Ces deux groupes de p -value sont ici représentés sous la forme d'une boîte à moustache pour chaque ensemble i (Fig 4.8) où la largeur de la boîte à moustache est fonction du nombre d'individus.

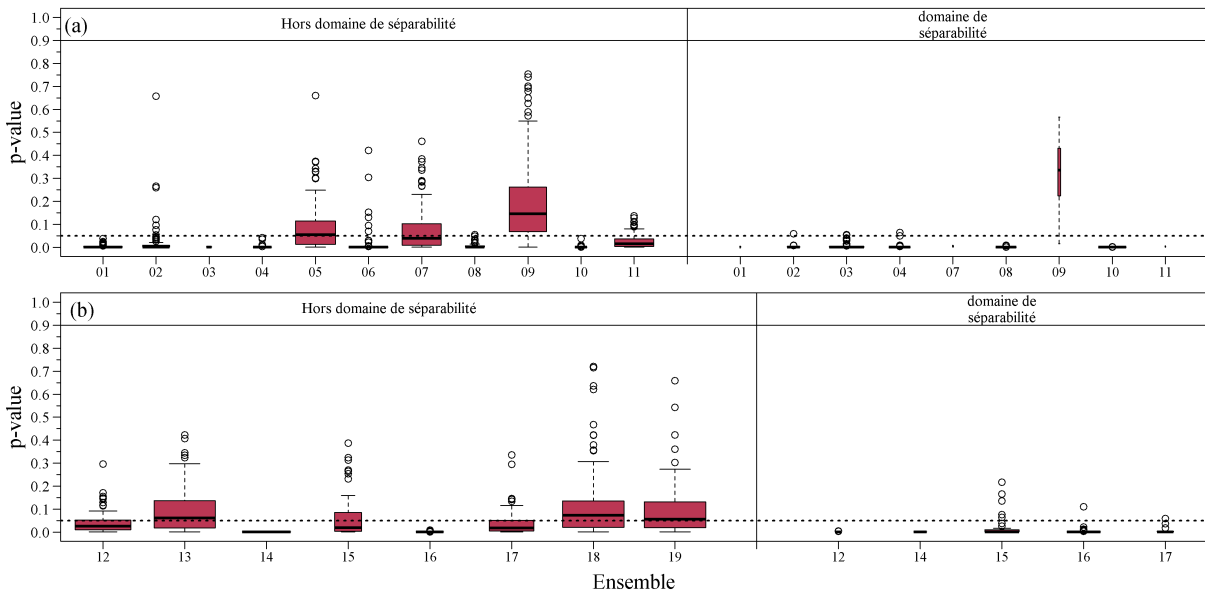


FIGURE 4.8 – Résultats de la comparaison entre les graphes de trajectoires indépendantes et les réalisations du modèle d'Erdős-Rényi. Pour chaque ensemble les p-values sont distinguées selon la position des graphes dans le domaine de séparabilité.

Cette comparaison entre les graphes de trajectoires indépendantes $G_{GAN}^{i,j}$ et des graphes d'arêtes indépendantes $G_{ER}(N^{i,j}, d^{i,j})$, conduit à des résultats très similaires de ceux observés lorsqu'il s'agissait de tester l'indépendance des relations dyadiques des graphes empiriques (Section 3.3.3). En effet d'une part, le test rejette l'hypothèse nulle H_0 pour la

majorité des graphes $G_{GAN}^{i,j}$ situés dans le domaine de séparabilité de la GCD_{11} . D'autre part, le test rejette également l'hypothèse nulle H_0 pour une large partie des graphes $G_{GAN}^{i,j}$ situés en dehors du domaine de séparabilité de la GCD_{11} . Ainsi au total, environ 76% (1453 sur 1900) des graphes de trajectoires indépendantes sont jugés significativement différents des réalisations du modèle d'Erdős-Rényi.

Ces résultats suggèrent sans ambiguïté qu'il est très improbable que des graphes issus de relations dyadiques entre trajectoires indépendantes présentent une organisation topologique similaire à celle attendue sous l'hypothèse d'indépendance des arêtes. Ces résultats permettent de réfuter l'hypothèse faisant le lien entre l'indépendance des trajectoires et celle des relations dyadiques, n'excluant donc pas la possibilité que les trajectoires empiriques, malgré leurs relations dyadiques non indépendantes, soient tout de même mutuellement indépendantes.

4.4 Graphes de trajectoires indépendantes *versus* graphes empiriques de flottilles

Même si les graphes empiriques et les graphes de trajectoires indépendantes sont tous deux différents des graphes d'indépendance des arêtes, il n'est pour autant pas possible d'affirmer qu'ils sont tous deux similaires. Dans cette dernière section, il est question d'interroger cette ultime similarité.

4.4.1 Comparaison du nombre de rencontres

Avant même de s'intéresser à la similarité topologique entre les graphes de trajectoires indépendantes et les graphes empiriques, une question préalable relative au nombre de relations dyadiques peut être posée. Un assemblage de trajectoires indépendantes, conduit-il à un nombre similaire de relations dyadiques que celui observé empiriquement ?

Une réponse à cette question peut être apportée par une comparaison entre la densité d'un graphe empirique G_{emp}^i et les densités des graphes de trajectoires indépendantes $G_{GAN}^{i,j}$ (Fig 4.9). Dans le cadre de cette comparaison, tous les graphes relatifs à un même ensemble i , qu'ils soient empiriques ou issus de trajectoires simulées, sont de même ordre N^i . Ainsi, la densité de ces graphes n'est fonction que du nombre de relations dyadiques (d'arêtes) dans le graphe.

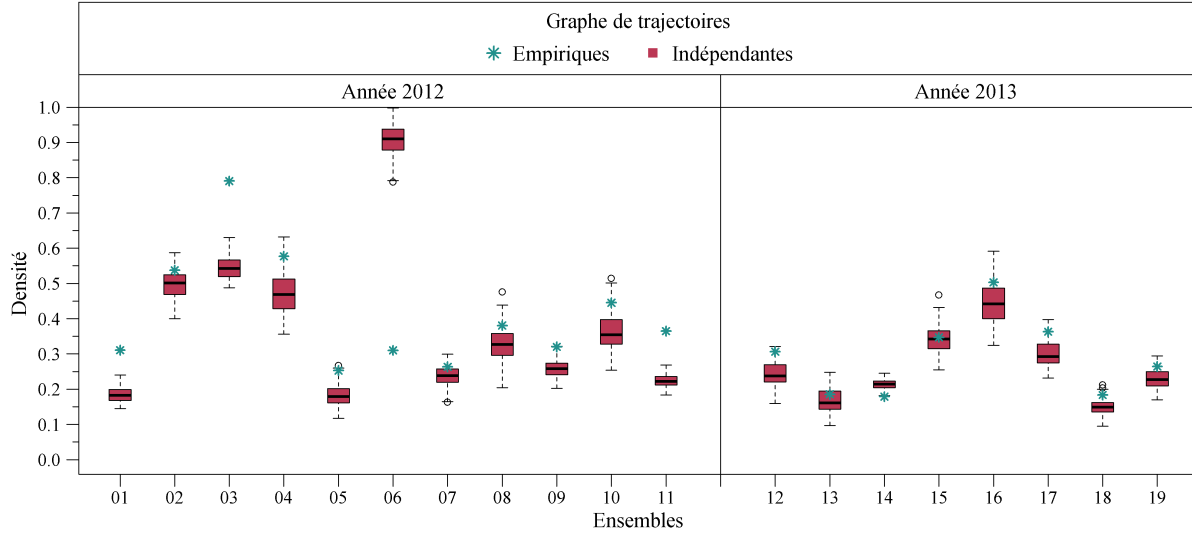


FIGURE 4.9 – Comparaison de la densité des graphes empiriques aux densités des graphes de trajectoires indépendantes.

Au regard de cette comparaison de la densité, il s'avère que d'une manière générale, les ensembles de trajectoires empiriques produisent un nombre de relations dyadiques supérieur à ceux observés dans les ensembles de trajectoires simulées. En effet, dans près de 80% des cas (15 des 19 ensembles i) la densité empirique des graphes G_{emp}^i est au moins supérieure au 3ème quartile de la distribution des densités des 100 graphes $G_{GAN}^{i,j}$. En d'autres termes, les trajectoires empiriques établissent en général plus de relations dyadiques que ce qui est observé en cas d'indépendance des trajectoires. Ces résultats sont en défaveur de l'hypothèse d'indépendance des trajectoires empiriques, néanmoins, ils doivent être corroborés avec la mise en évidence d'une réelle différence d'organisation topologique arêtes.

4.4.2 Comparaison des graphes empiriques aux graphes des trajectoires indépendantes

Considérant un ensemble i , défini par un graphe empirique G_{emp}^i et une collection de 100 graphes de trajectoires indépendantes $G_{GAN}^{i,j}$, le test statistique développé dans ce travail peut être utilisé afin d'évaluer la probabilité que ce graphe empirique G_{emp}^i appartienne à la collection des graphes $G_{GAN}^{i,j}$. En d'autres termes, il s'agit ici de substituer les réalisations du modèle nul, préalablement incarné par le modèle d'Erdős-Rényi, par la collection de graphes $G_{GAN}^{i,j}$.

A la différence de toutes les comparaisons réalisées précédemment, il s'agit ici de comparer un graphe empirique d'ordre N^i et de densité d^i , avec une collection de graphes de même ordre N^i mais dont la densité $d^{i,j}$ peut être différente. Ainsi, de la même manière qu'il a été nécessaire de s'intéresser au domaine de séparabilité de la GCD_{11} lors de la comparaison à ordre et à densité identique, cette fois-ci, il est nécessaire de s'intéresser au domaine de séparabilité obtenu lors de la comparaison à ordre identique mais à densité différente. Comme cela a été montré au cours du chapitre précédent, dans cette configuration de comparaison et pour les gammes d'ordres en présence ici (entre 30 et 45), les performances de la GCD_{11} ne sont pas suffisantes pour parvenir à produire des résultats satisfaisants.

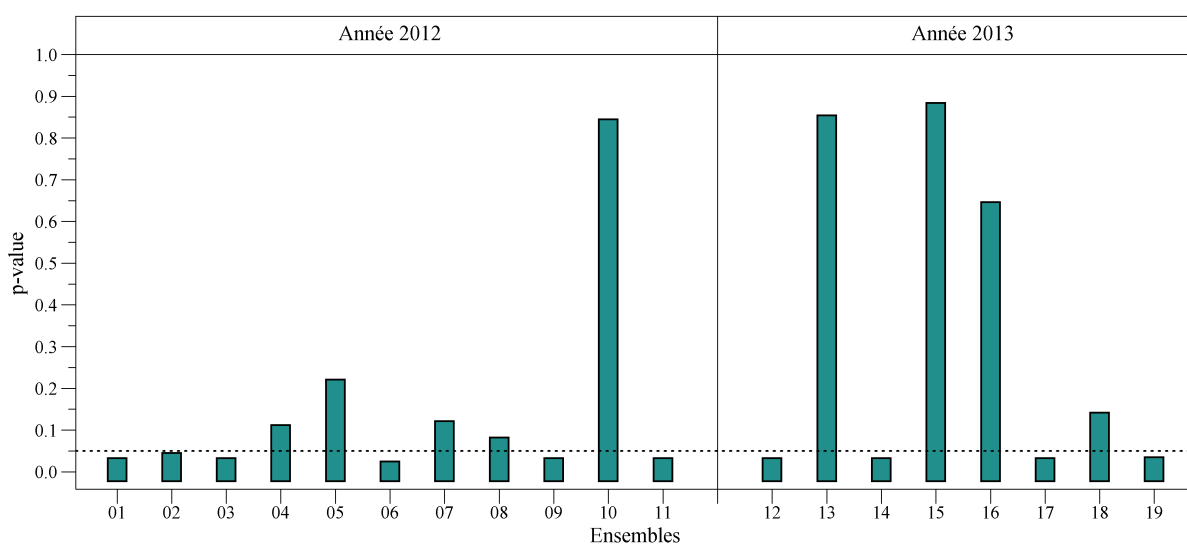


FIGURE 4.10 – Résultat des tests statistiques issus de la comparaison entre les graphes empiriques et les graphes de trajectoires indépendantes. Dans cette configuration, la p – $value$ renseigne la probabilité qu'un graphe empirique G_{emp}^i soit similaire aux 100 graphes de trajectoires indépendantes $G_{GAN}^{i,j}$

Le test statistique peut tout de même être réalisé entre les graphes empiriques et les graphes de trajectoires indépendantes (Fig 4.10). Ainsi parmi les 19 graphes empiriques G_{emp}^i de la flottille OTB, 52% (10 sur 19) d'entre eux sont jugés statistiquement différents (p – $value < 0.05$) de leur ensemble de 100 graphes de trajectoires indépendantes $G_{GAN}^{i,j}$. A l'inverse, 48% (9 sur 19) des graphes empiriques sont jugés similaires à leur ensemble de graphes de trajectoires indépendantes, dont 4 d'entre eux, l'ensemble $\{10, 13, 15$ et $16\}$, présentent une très forte p – $value$. Malgré ces fortes p – $values$, les faibles ordres de ces 4 graphes $\{31, 32, 30, 31\}$, les plaçant dans un domaine où les performances de la GCD_{11}

ne sont pas assurées, ne permettent pas de conclure en faveur d'une réelle similarité entre les graphes empiriques et les graphes des trajectoires indépendantes.

4.5 Conclusion du chapitre

Dans ce dernier chapitre nous avons totalement réformé notre approche visant à interroger l'indépendance des trajectoires en substituant les modèles de graphes aléatoires, socle du chapitre précédent, par des modèles de simulation de trajectoires, représentés par les GAN. Grâce à ces modèles, nous avons pu reconstruire des ensembles de trajectoires indépendantes et réalistes au regards de ceux observés empiriquement. Sur la base de ces ensembles de trajectoires indépendantes, nous avons pu tester et réfuter l'hypothèse de travail du chapitre précédent liant l'indépendance des relations dyadiques à celle des trajectoires.

Finalement, une dernière série de comparaisons entre les graphes empiriques et les graphes de trajectoires indépendantes nous a placé dans le cadre d'une comparaison de graphes d'ordres N et de densités d différents. Ces derniers résultats nous permettent de réfuter l'hypothèse d'indépendance pour une moitié des ensembles de trajectoires de la flottille OTB tandis que pour l'autre moitié, leur indépendance n'est pas à exclure.

DISCUSSION GÉNÉRALE

Dans ce manuscrit, notre objectif était de fournir une approche permettant d'analyser l'organisation collective d'un groupe de navires de pêche. La réalisation de cet objectif s'est traduite par le développement de deux approches focalisées sur l'identification d'une propriété particulière, l'indépendance mutuelle des navires de pêche, ou plus précisément celle de leurs trajectoires. En effet, nous considérons ici que démontrer l'indépendance des trajectoires est d'une certaine manière synonyme de démontrer une absence de comportement collectif. Ce travail nous a conduit à nous intéresser d'une part, à la manière de comparer des schémas d'organisation d'un groupe de navires, et d'autre part, à identifier ce qui, dans les schémas d'organisation, traduit l'indépendance. A cette fin, deux groupes de navires regroupés en flottille ont été considérés, des chalutiers en boeufs PTM (flottille de référence de la non indépendance deux à deux), et des chalutiers de fond OTB (flottille dont l'indépendance est à tester).

S'intéresser à l'organisation collective d'un groupe de navires revient à s'intéresser à la manière dont les navires s'organisent les uns par rapport aux autres lorsqu'ils sont en mer. C'est sur cette considération que s'est ouvert le chapitre 2 présentant une approche s'intéressant au niveau d'expression le plus bas de l'organisation collective, à savoir, l'organisation deux à deux appréhendée au travers des dyades. Reposant sur un critère de proximité physique simultanée, l'identification des dyades permet, dans un premier temps, de décrire l'organisation spatiale et temporelle de chaque couple de trajectoires sous la forme d'une question binaire : "ces deux navires ont-ils été oui ou non au moins une fois au même moment au même endroit?". Dans un second temps, l'intégration de l'ensemble des dyades sous la forme d'un graphe, a permis un changement d'échelle en mettant en perspective l'organisation spatiale et temporelle des trajectoires à l'échelle d'un ensemble de trajectoires d'une flottille. Sous cette forme de graphe, la question posée est plus large : "au sein d'une flottille, quels sont les navires ayant été au moins une fois au même moment au même endroit?" En d'autres termes, le graphe des relations dyadiques fournit un résumé non exhaustif de l'information spatiale et temporelle contenue dans chaque trajectoire.

Au cours de ce chapitre nous avons également montré que l'analyse des dyades par

l'utilisation de métriques descriptives et de méthodes de classification permet d'interpréter les dyades comme la traduction de différents types de comportements. Ainsi nous avons montré au travers de son application à la flottille PTM, que cette approche était pertinente notamment pour identifier les couples de trajectoires traduisant de très fort appariement (collaborations "de type PTM", dyades à forte cohésion, groupe n°1). Toutefois, l'approche par classification et le choix d'une discrimination en trois groupes a fourni des résultats peu satisfaisants en ne parvenant pas à clairement discriminer les dyades de faible cohésion (groupe n°3) et de moyenne cohésion (groupe n°2).

Le chapitre 3 nous a amené à nous interroger sur la manière de comparer des schémas d'organisation des arêtes d'un graphe et d'évaluer la similarité entre deux graphes. Le choix de cette approche a reposé sur le fait que certains modèles de graphes produisent des réalisations présentant des propriétés d'organisations d'arêtes connues. Ce chapitre a ainsi été dans un premier temps le lieu du développement d'un test statistique reposant sur la distance GCD_{11} permettant d'évaluer la probabilité qu'un graphe empirique appartienne à une collection de graphes. Dans un second temps, ce test a pu être appliqué uniquement sur des ensembles restreints de graphes de flottille PTM et OTB afin d'évaluer leurs similarités respectives à des réalisations de modèles d'indépendance des arêtes Erdős-Rényi. Nous avons ainsi montré qu'il était très vraisemblable que les arêtes de ces graphes empiriques, c'est-à-dire les relations dyadiques, ne sont pas mutuellement indépendantes.

Finalement, une dernière réflexion sur l'hypothèse qu'il existe un lien entre l'indépendance des trajectoires et l'indépendance des relations dyadiques nous a conduits au cours du chapitre 4 à nous intéresser aux approches de simulation de trajectoires. Ainsi à l'aide de modèle GAN, nous avons pu simuler des ensembles de trajectoires réalistes au regard des trajectoires empiriques, et nous assurer que les trajectoires simulées étaient mutuellement indépendantes. Nous avons alors montré dans un premier temps que des ensembles de trajectoires indépendantes ne se traduisent pas par des graphes aux arêtes mutuellement indépendantes, invalidant l'hypothèse d'une relation de causalité entre l'indépendance des trajectoires et celle des relations dyadiques. Dans un second temps nous avons finalement conclu qu'une partie des graphes empiriques de la flottille OTB ne présentait pas une organisation des arêtes similaire à celle observée sur des ensembles de trajectoires indépendantes. Toutefois on ne peut pas exclure que l'autre partie de ces graphes de la flottille OTB traduise des trajectoires indépendantes.

Derrière la question de l'indépendance des trajectoires posée dans ce manuscrit, se

trouve la question de l'indépendance des traits de chaluts et des captures réalisées entre les différents navires, en d'autres termes l'indépendance des "échantillonnages" effectués par les pêcheurs. Cette question nous a conduit à interroger l'organisation des trajectoires les unes par rapport aux autres au travers de leurs relations dyadiques. Toutefois en conditionnant l'existence d'une dyade par un unique critère de proximité simultanée (5 km), cette approche nous a amené à interroger d'une manière assez peu supervisée l'organisation de ces trajectoires. En effet, sur la base de ce critère, une partie des dyades identifiées est relative à des aspects de l'organisation des trajectoires qui ne semblent pas en lien direct avec l'action de pêche, comme par exemple deux navires se croisant, ou quittant simplement le même port. Il pourrait être envisageable de conditionner l'existence d'une dyade à des critères plus restreints et plus en lien avec l'activité halieutique, comme par exemple, inclure en plus du critère de proximité, un critère relatif à la vitesse des navires avec un seuil fixé par exemple à 4.5 noeuds. En considérant qu'une vitesse inférieure à ce seuil traduit une action de pêche, les graphes construits à partir de dyades respectant ces critères pourraient fournir une représentation de l'organisation des navires ayant exploités simultanément la même zone de pêche, et n'incluraient pas les dyades issues du déplacement de types croisement, ou entrée et sortie simultanées du port. Ainsi, dans un premier temps, les trajectoires identifiées comme n'ayant formé aucune dyade pourraient être considérées comme indépendantes (noeuds isolés). Dans un second temps, une analyse des dyades au travers de leurs métriques descriptives pourrait permettre de différencier, au travers de l'évaluation de leur cohésion, les couples de trajectoires traduisant d'une rencontre ponctuelle et fortuite sur une même zone de pêche, de ceux traduisant des rencontres régulières et répétées.

Le critère d'existence d'une dyade reposant sur la simultanéité des trajectoires pourrait également être modifié pour de futures analyses. En effet, un autre aspect de l'organisation des trajectoires pourrait être interrogé en s'intéressant à sa dynamique dans le temps. Un navire ayant réalisé une marée particulièrement intéressante au regard de ses objectifs de capture, pourrait être amené à réitérer une marée similaire afin d'exploiter les mêmes zones de pêches que durant sa marée précédente. En autorisant l'établissement d'une dyade entre une marée et celles du passé, deux marées d'un même navire pourraient constituer des dyades à forte cohésion traduisant une forme de dépendance d'un navire à lui même ou plus largement des dépendances dans le temps entre des ensembles de trajectoires.

C'est toutefois sur des dyades définies grossièrement qu'une première approche centrée sur la comparaison des graphes de flottille et des réalisations du modèle d'Erdős-Rényi

a été développée ici. A cette fin, nous avons proposé un test statistique reposant sur la distance GCD_{11} afin d'évaluer la similarité entre des ensembles de graphes. Afin de rendre pertinente la comparaison des graphes de flottilles aux réalisations du modèle d'Erdős-Rényi, il a été nécessaire de réduire spatialement et temporellement le cas d'étude, se limitant alors aux activités hebdomadaires des ports de Cherbourg pour la flottille PTM et de Boulogne-sur-Mer pour la flottille OTB. Cette réduction nous a conduit à étudier les performances de la GCD_{11} lors de la comparaison de graphes d'ordres faibles (inférieurs à 50 noeuds) et à délimiter le domaine de séparabilité de la GCD_{11} . Pour cette application, l'adaptabilité de la Graphlet Correlation Distance au choix des orbites utilisées pour évaluer ses performances et pour réaliser les comparaisons au modèle d'Erdős-Rényi n'a pas été évoquée. En effet, les 11 orbites définissant la GCD_{11} initialement jugées pertinentes par Yaveroglu et al, l'ont été sur la base de comparaisons de graphes se tenant dans des gammes d'ordres et de densités restreintes ($1000 \geq N \geq 6000$ et $0.05 \geq d \geq 0.1$). Dans ce travail, notre contribution a consisté à explorer de nouvelles gammes d'ordres et de densités ($5 \geq N \geq 100$ et $0 \geq d \geq 1$) mais n'a pas exploré les effets d'un changement de combinaison d'orbites dans les performances de la GCD . Il est envisageable que certaines combinaisons d'orbites soient plus pertinentes que d'autres en fonction des gammes d'ordres et de densités et selon le type de comparaisons à effectuer, à ordres et densités identiques ou différents. Si tel est le cas, le domaine de séparabilité pourrait être fonction de différentes combinaisons d'orbites et de meilleures performances pourraient être observées lors de la comparaison des petits graphes.

De plus, nous avons montré lors de la comparaison des graphes de flottilles aux réalisations du modèle d'Erdős-Rényi, que le rejet de l'hypothèse nulle peut être dû à des différences significatives d'un faible nombre de coefficients de corrélation des Graphlets impliquant seulement quelques orbites. Le recours au test statistique pourrait ainsi être prolongé par l'identification des coefficients de corrélation responsables du rejet de l'hypothèse nulle. Une analyse de ces coefficients pourrait alors renseigner sur quels aspects topologiques les graphes diffèrent.

Finalement le dernier chapitre, le chapitre 4, est apparu comme une réponse aux questionnements des chapitres précédents. Tout d'abord, l'approche par simulation au travers des modèles de GAN en produisant des trajectoires indépendantes mais réalistes, permet de produire des graphes dont l'organisation topologique des dyades rend compte des contraintes spatiales et temporelles des trajectoires (géométrie des zones de pêche au cours du temps). En d'autres termes, les graphes des trajectoires simulées traduisent

d'une certaine manière l'hypothèse nulle à savoir l'organisation des relations dyadiques de trajectoires indépendantes conditionnellement au fait que les trajectoires proviennent d'un même port et partagent un même espace spatial et temporel. De plus, le fait de pouvoir simuler des ensembles de trajectoires permet d'appliquer à ces ensembles la même méthode de construction de graphes que celle appliquée aux trajectoires empiriques, autorisant ainsi une certaine liberté quant au choix des critères d'existences d'une dyade. En effet, même si par exemple un changement du critère de proximité conduit à une modification du graphe empirique, on s'attend à ce que, une modification similaire se produise sur les graphes des trajectoires simulées sans modification sur les résultats du test statistique.

Dans ce travail, la notion d'indépendance a été abordée sous un angle binaire à un ensemble de trajectoires de sorte à produire un diagnostic de type l'ensemble est indépendant ou il ne l'est pas. Ceci est appuyé par un test statistique de rejet ou non de l'hypothèse nulle. Toutefois, une certaine nuance pourrait être apportée. Supposons un ensemble de 98 trajectoires indépendantes dans lequel on intègre 2 trajectoires en forte interaction (typiquement deux trajectoires de navires PTM formant une dyade à forte cohésion), et comparons le graphe des relations dyadiques de cet ensemble de 100 trajectoires à des graphes de trajectoires indépendantes. Dans cette situation, on s'attendrait à ce que le test statistique rejette l'hypothèse nulle (les trajectoires de cet ensemble sont jugées non indépendantes car 98% d'entre elles le sont). Il est très peu probable qu'une telle situation existe empiriquement, cependant il est probable qu'au sein d'un ensemble de trajectoires, certaines d'entre elles soient plus indépendantes que d'autres. Dans ce cas l'évaluation de la cohésion des dyades pourrait être utile pour identifier les couples de trajectoires les moins indépendants deux à deux et étudier l'effet de leur suppression sur l'indépendance mutuelle du reste de l'ensemble.

En conclusion, le travail développé dans ce manuscrit a consisté à mobiliser et à faire interagir des concepts et des méthodes issus de différents domaines scientifiques, allant des méthodes d'analyse de trajectoires aux approches d'apprentissage automatique et profond, le tout articulé autour de la théorie des graphes. En cherchant à développer une approche permettant d'interroger l'organisation d'un ensemble de trajectoires, ce travail a pris une tournure méthodologique conduisant à la réduction du cadre d'étude de l'échelle de la Manche à l'échelle de deux ports, réduisant également la portée des résultats obtenus. En d'autres termes, ce travail ne permet pas de répondre formellement à la question initialement posée sur l'indépendance des trajectoires d'une flottille. Toutefois, les différentes pistes abordées tout au long de ce manuscrit devraient permettre de fournir

des éléments clefs d'une application approfondie, notamment l'approche par simulation à l'aide de modèle GAN qui pourrait permettre de revenir au cadre initial de l'étude à l'échelle de plusieurs ports simultanément, voire de la Manche entière.

Le point bleu pâle

Notre planète est une poussière isolée, enveloppée dans la grande nuit cosmique. Dans notre obscurité, dans toute cette immensité, rien ne laisse présager qu'une aide viendra d'ailleurs, pour nous sauver de nous-mêmes. La Terre est jusqu'à présent le seul monde connu à abriter la vie. Il n'y a nulle part ailleurs, au moins dans un futur proche, vers où notre espèce pourrait migrer. Visiter, oui. S'installer, pas encore. Que vous le vouliez ou non, pour le moment, c'est sur Terre que nous nous trouvons.

(C.Sagan [169])

INFORMATIONS SUPPLÉMENTAIRES

Supporting information

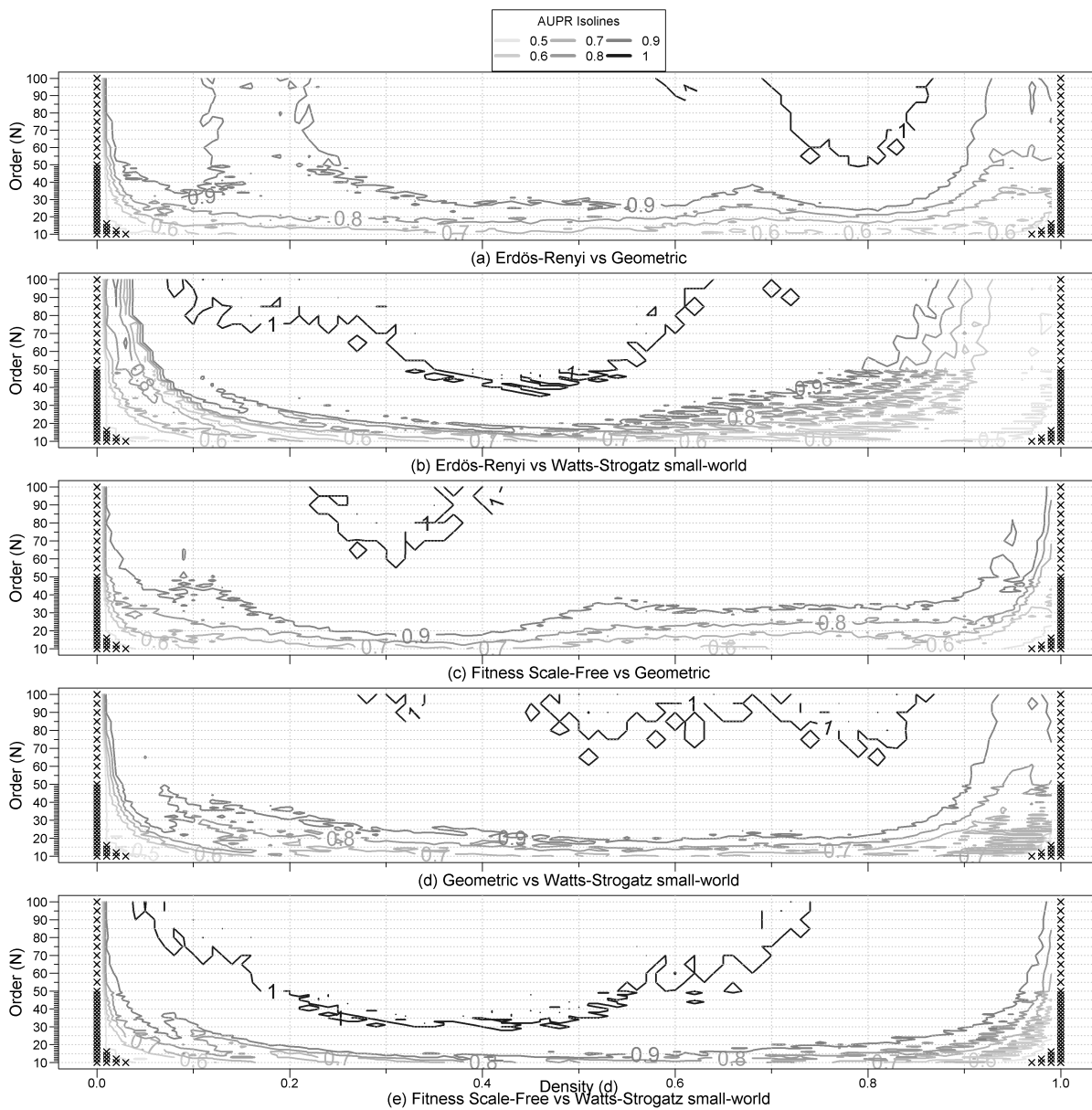
S1 Eq To account for the difference in variability between the correlation coefficients of each pair of orbits, we also computed the following standardised distance δ_{std,M_k} between $GCM(M_k)$ and \overline{GCM}_M where $\sigma(i, j)$ is the standard deviation of the correlation coefficients of the pair of orbits (i, j) under H_0 . We built the test by computing η the number of times the standardised distance between GCM_G and \overline{GCM}_M is smaller or equal to the distance δ_{std,M_k} . The p -value [50] is defined by $\hat{p} = (\eta + 1)/(K + 1)$. The larger the p -value, the less evidence against H_0 .

$$\delta_{std,M_k} = \sqrt{\sum_{i=1}^{11} \sum_{j=i+1}^{11} \left(\frac{\overline{GCM}_M(i, j) - GCM(M_k)(i, j)}{\sigma(i, j)} \right)^2} \quad (4.1)$$

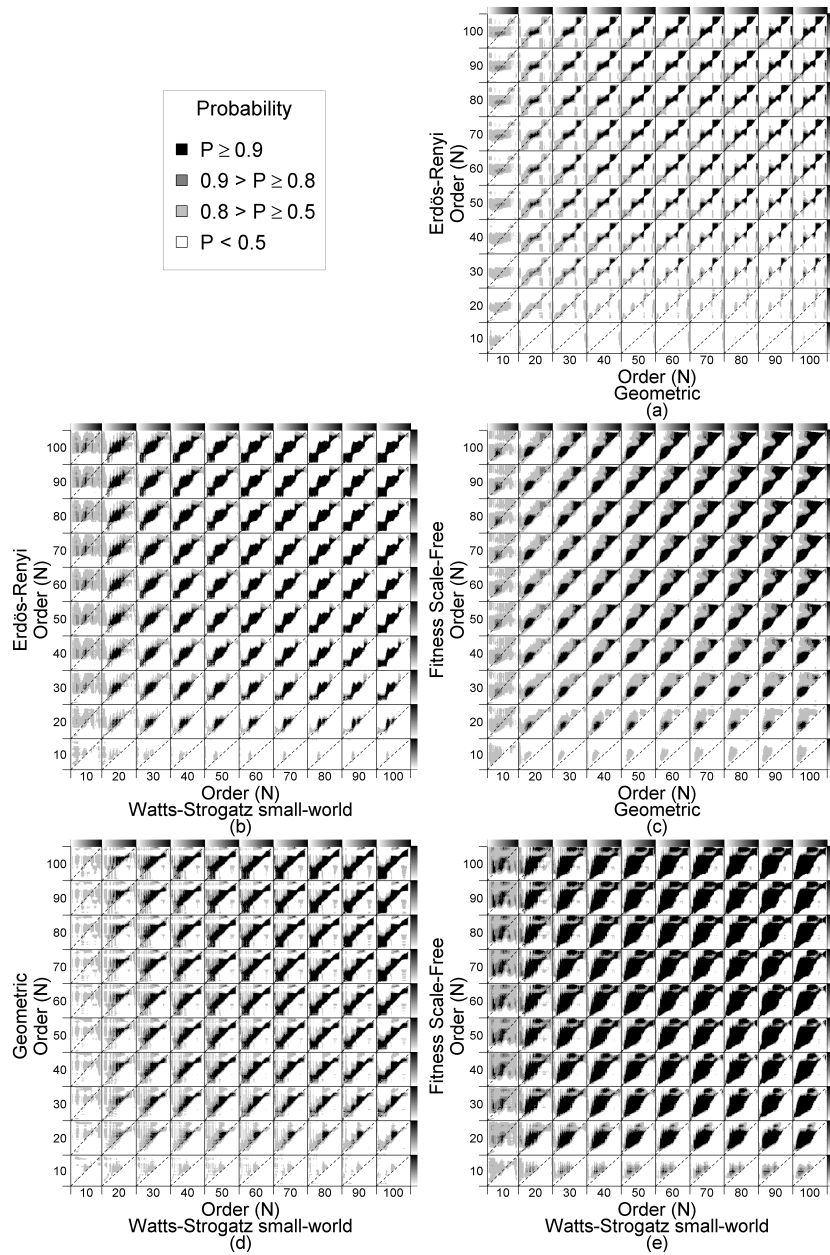
S1 Table Estimated p-values (std). Each empirical graph is associated with an estimated p -value (\hat{p}) of being an outcome of an Erdős-Rényi, Fitness scale-free model, a Watts-Strogatz small world or a Geometric model. As in Table 3.3, empirical graphs are sorted according to their order. ($\hat{p}^* < 0.05$, $\hat{p}^{**} < 0.01$ and $\hat{p}^{***} \leq 0.001$)

Graph	Standardised P value			
	Erdős Renyi	Scale Free	Watts-Strogatz	Geometric
Graph 1	0.001***	0.001***	0.001***	0.095
Graph 2	0.001***	0.001***	0.001***	0.160
Graph 3	0.001***	0.001***	0.002**	0.092
Graph 4	0.001***	0.020*	0.001***	0.039*
Graph 5	0.001***	0.001***	0.001***	0.102
Graph 6	0.001***	0.001***	0.001***	0.001***
Graph 7	0.001***	0.001***	0.001***	0.003**
Graph 8	0.001***	0.001***	0.001***	0.004**
Graph 9	0.001***	0.001***	0.001***	0.003**
Graph 10	0.001***	0.001***	0.001***	0.173

S1 Fig Quality of clustering (AUPR) for pairs of models. Diagram of the matrix of AUPR (A_{M_1, M_2}) (a) Erdős-Rényi vs Geometric (b) Erdős-Rényi vs Watts-Strogatz small-world (c) Fitness Scale-Free vs Geometric (d) Geometric vs Watts-Strogatz small-world (e) Fitness Scale-Free vs Watts-Strogatz small-world. For each pair of models, and for each order (from 10 to 100) and edge density (from 0 to 1) combination, the quality of clustering between 100 graphs of each of the two types of models is assessed by the Area Under the Precision-Recall curve (AUPR). A maximum value of 1 corresponds to perfect discrimination. Black crosses represent zero AUPR.



S2 Fig Probability of correctly distinguishing two random graph models with different order and/or edge density. Each block (i, j) concerns the comparison of an M_1 of order N and a M_2 of order N' , with edge density d and d' respectively ranging from 0 to 1 (grey gradient from white to black on the top and right side) with $(M_1, M_2) \in \{ER, SF, SW, GO\}^2$ and $M_1 \neq M_2$. Dashed lines in each block highlight comparison when $d = d'$. (a) Probability matrix B_{M_1, M_2} that $D_{M_1(N, d), M_2(N', d')} > \max(D_{M_1(N, d), M_1(N', d')}, D_{M_2(N, d), M_2(N', d')})$ with $M_1 = SF$ and $M_2 = GO$. (b) $B_{ER, SW}$, (c) $B_{SF, GO}$, (d) $B_{GO, SW}$, (e) $B_{SF, SW}$.



Bibliographie

1. SCHOPENHAUER, A., *Le monde comme volonté et comme représentation* (Brockhaus, Leipzig, 1819).
2. CAMUS, A., *Le mythe de Sisyphe* (Good Press, 2021).
3. UNIES, N., Report of the World Summit on Sustainable Development, *Johannesburg, South Africa* (2002).
4. UNIES, N., Protocole de Kyoto la convention-cadre des Nations Unies sur les changements climatiques, *Kyoto, Japon* (1998).
5. MEADOWS, D. H., MEADOWS, D. H., RANDERS, J. & BEHRENS III, W. W., The limits to growth : a report to the club of Rome (1972), *Google Scholar* **91**, 2 (1972).
6. CHANGE, C. *et al.*, Intergovernmental panel on climate change, *World Meteorological Organization* **52** (2007).
7. HEINBERG, R., *Peak everything : waking up to the century of declines* (New Society Publishers, 2010).
8. CAPUANO, L., International energy outlook 2018 (IEO2018), *US Energy Information Administration (EIA) : Washington, DC, USA* **2018**, 21 (2018).
9. BLAMONT, J., *Introduction au siècle des menaces* (Odile Jacob, 2004).
10. STEWART, I., *Does God play dice ? : The new mathematics of chaos* (Penguin UK, 1997).
11. BAUDELAIRE, C., *Les fleurs du mal* (Auguste Poulet-Malassis, Paris, 1857).
12. HOMÈRE, *Odyssée* (Auguste Poulet-Malassis, Grèce antique, 5eme av. J.C).
13. VIGNAUD, H., *Le vrai Christophe Colomb et la légende : la date exacte de la naissance du grand Génois, sa famille, les indications qu'il avait, Toscanelli, prétendu initiateur de la découverte de l'Amérique, l'objet véritable de l'entreprise de 1492* (Auguste Picard, Paris, 1921).
14. DE BOUGAINVILLE, L.-A., *Voyage autour du monde* (Presses de l'Université de Paris-Sorbonne, Paris, 1771).
15. GALVÃO, A., HAKLUYT, R. & BETHUNE, C. R. D., *The discoveries of the world, from their first original unto the year of Our Lord 1555* (Printed for the Hakluyt society, Londres, 1862).

-
16. APPLEYARD, S. F., *Marine electronic navigation* (Routledge, 2006).
 17. MCNEFF, J. G., The global positioning system, *IEEE Transactions on Microwave theory and techniques* **50**, 645-652 (2002).
 18. LORD, W., *The Night Lives On : The Untold Stories and Secrets Behind the Sinking of the " Unsinkable " Ship—Titanic* (Open Road Media, 2012).
 19. FREMONT, A., Le transport maritime depuis 1945 : facteur clé de la mondialisation, *Entreprises et histoire*, 16-29 (2019).
 20. COUTANSAIS, C. P., Transports. Transport maritime : entre globalisation et développement durable—Études, *Problemes Economiques*, 34 (2011).
 21. HOFFMANN, J. & SIRIMANNE, S., in *United Nations Conference on Trade and Development* (2017).
 22. FOURASTIÉ, J., *Les trente glorieuses : ou la Révolution invisible de 1946 à 1975* (Fayard, 2014).
 23. BERNARD, N., Du port-parking au produit touristique : l'évolution des ports de plaisance en France, *Norois* **182**, 275-285 (1999).
 24. ROUSSEAU, Y., WATSON, R. A., BLANCHARD, J. L. & FULTON, E. A., Evolution of global marine fishing fleets and the response of fished resources, *Proceedings of the National Academy of Sciences* **116**, 12238-12243 (2019).
 25. FELSKI, A., JASKÓLSKI, K. & BANYŚ, P., Comprehensive assessment of automatic identification system (AIS) data application to anti-collision manoeuvring, *The Journal of Navigation* **68**, 697-717 (2015).
 26. OMER, M., MOSTASHARI, A., NILCHIANI, R. & MANSOURI, M., A framework for assessing resiliency of maritime transportation systems, *Maritime Policy & Management* **39**, 685-703 (2012).
 27. HU, Y. & ZHU, D., Empirical analysis of the worldwide maritime transportation network, *Physica A : Statistical Mechanics and its Applications* **388**, 2061-2071 (2009).
 28. LEE, J. M.-y. & WONG, E. Y.-c., *Suez Canal blockage : an analysis of legal impact, risks and liabilities to the global supply chain* in *MATEC Web of Conferences* **339** (2021).

-
29. CAMPANA, I. *et al.*, Seasonal characterisation of maritime traffic and the relationship with cetacean presence in the Western Mediterranean Sea, *Marine Pollution Bulletin* **115**, 282-291 (2017).
 30. FILICOTTO, F. *et al.*, Behavioural and biochemical stress responses of *Palinurus elephas* after exposure to boat noise pollution in tank, *Marine pollution bulletin* **84**, 104-114 (2014).
 31. MILAZZO, M., BADALAMENTI, F., CECCHERELLI, G. & CHEMELLO, R., Boat anchoring on *Posidonia oceanica* beds in a marine protected area (Italy, western Mediterranean) : effect of anchor types in different anchoring stages, *Journal of Experimental Marine Biology and Ecology* **299**, 51-62 (2004).
 32. NEPOTE, E., BIANCHI, C. N., MORRI, C., FERRARI, M. & MONTEFALCONE, M., Impact of a harbour construction on the benthic community of two shallow marine caves, *Marine pollution bulletin* **114**, 35-45 (2017).
 33. KIVEKÄS, N. *et al.*, Contribution of ship traffic to aerosol particle concentrations downwind of a major shipping lane, *Atmospheric Chemistry and Physics* **14**, 8255-8267 (2014).
 34. SHEPHERD, J., Oceans of truth, *Nature* **388**, 843-843 (1997).
 35. PARLIAMENT, U., NORTH SEA (OVER-FISHING), *Historic Hansard* **438** (1947).
 36. FAO, F. *et al.*, Food and agriculture organization of the United Nations, Rome, URL : <http://faostat.fao.org> (2018).
 37. FAO., A., The State of World Fisheries and Aquaculture 2020 (Sustainability in Action), Rome (2020).
 38. HUTCHINGS, J. A. & MYERS, R. A., What can be learned from the collapse of a renewable resource? Atlantic cod, *Gadus morhua*, of Newfoundland and Labrador, *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 2126-2146 (1994).
 39. ROBERTS, J. M., WHEELER, A. J. & FREIWALD, A., Reefs of the deep : the biology and geology of cold-water coral ecosystems, *Science* **312**, 543-547 (2006).
 40. CONTI, A., *Racleurs d'océans* (Éd. Payot & Rivages, 2002).
 41. SYVERSEN, T., LILLENG, G., VOLLSTAD, J., HANSSSEN, B. J. & SØNVISEN, S. A., Oceanic plastic pollution caused by Danish seine fishing in Norway, *Marine Pollution Bulletin* **179**, 113711 (2022).

-
42. ROUSSEAU, J.-J. *et al.*, Discours sur l'origine et les fondements de l'inégalité parmi les hommes, suivi de La reine fantasque, *Discours sur l'origine et les fondements de l'inégalité parmi les hommes, suivi de La reine fantasque*, 1-190 (2009).
 43. NIELSEN, L. A., The evolution of fisheries management philosophy, *Marine Fisheries Review* **38**, 15-23 (1976).
 44. JACQUET, J., Silent water : a brief examination of the marine fisheries crisis, *Environment, Development and Sustainability* **11**, 255-263 (2009).
 45. ALLISON, E. H., Big laws, small catches : global ocean governance and the fisheries crisis, *Journal of International Development* **13**, 933-950 (2001).
 46. NATIONS, U., *Report of the International Technical Conference on the Conservation of the Living Resources of the Sea (18 April to 10 May 1955, Rome)*. (UN, 1955).
 47. WHITEMAN, M. M., Conference on the Law of the Sea : Convention on the Continental Shelf, *American Journal of International Law* **52**, 629-659 (1958).
 48. NATIONS, U., *Report of the United Nations Conference on the Human Environment : Stockholm, 5-16 June 1972* (1973).
 49. BRUNDTLAND, G. H., *Report of the World Commission on environment and development : " our common future. "* (UN, 1987).
 50. RUSSELL, E. S., Some theoretical considerations on the "overfishing" problem, *ICES Journal of Marine Science* **6**, 3-20 (1931).
 51. UNIES, N., Rapport du Sommet mondial pour le développement durable, *Johannesburg (Afrique du Sud)* **26**, 9 (2002).
 52. FRAMEWORK, K.-M. G. B., *Draft Decision Submitted by the President* in (2022).
 53. EUROPÉEN, P., La Politique Commune de la Pêche : Origines et Évolution, *europarl. eurpa* (2023).
 54. DEPERROIS, H. & LAISNE, L., Mission sur la politique commune des pêches, Rapport CGAAER n°21019, IGAM n°2021-034, *Inspection Générale des Affaires Maritimes* (2021).
 55. LACROIX, A., LAURENT, F., RUELLAND, D. & SAUBOUA, E., *Diagnostic de la pollution nitrique : du modèle à l'indicateur* in *Symposium International INRA-PSDR, " Territoires et enjeux du développement régional "*, Lyon (2005), 9-11.

-
56. GARCIA, S., Indicators for sustainable development of fisheries, *FAO Land and Water Bulletin (FAO)* (1997).
 57. NISHIDA, T., UEHARA, S. & NYUNDO, R., Predatory behavior among wild chimpanzees of the Mahale Mountains, *Primates* **20**, 1-20 (1979).
 58. ROCHET, M.-J., BISEAU, A., CORNOU, A.-S., DELAUNAY, D. & SCAVINNER, M., Utilisation des données Obsmer–bilan 2009-2015 (2017).
 59. PATTERSON, T. A., THOMAS, L., WILCOX, C., OVASKAINEN, O. & MATTHIOPOULOS, J., State–space models of individual animal movement, *Trends in ecology & evolution* **23**, 87-94 (2008).
 60. JONSEN, I. D., MYERS, R. A. & JAMES, M. C., Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model, *Marine Ecology Progress Series* **337**, 255-264 (2007).
 61. NATALE, F., GIBIN, M., ALESSANDRINI, A., VESPE, M. & PAULRUD, A., Mapping fishing effort through AIS data, *PloS one* **10**, e0130746 (2015).
 62. VERMARD, Y., RIVOT, E., MAHÉVAS, S., MARCHAL, P. & GASCUEL, D., Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models, *Ecological Modelling* **221**, 1757-1769 (2010).
 63. GLOAGUEN, P. *et al.*, An autoregressive model to describe fishing vessel movement and activity, *Environmetrics* **26**, 17-28 (2015).
 64. ALGLAVE, B. *et al.*, Combining scientific survey and commercial catch data to map fish distribution, *ICES Journal of Marine Science* **79**, 1133-1149 (2022).
 65. JOO, R. *et al.*, Identifying partners at sea from joint movement metrics of pelagic pair trawlers, *ICES Journal of Marine Science* **78**, 1758-1768 (2021).
 66. TUTTE, W. T. & TUTTE, W. T., *Graph theory* (Cambridge university press, 2001).
 67. WEST, D. B. *et al.*, *Introduction to graph theory* (Prentice hall Upper Saddle River, 2001).
 68. DERRIDA, B., MANRUBIA, S. C. & ZANETTE, D. H., Statistical properties of genealogical trees, *Physical Review Letters* **82**, 1987 (1999).
 69. HOLTEN, D., Hierarchical edge bundles : Visualization of adjacency relations in hierarchical data, *IEEE Transactions on visualization and computer graphics* **12**, 741-748 (2006).

-
70. PIMM, S. L., LAWTON, J. H. & COHEN, J. E., Food web patterns and their consequences, *Nature* **350**, 669-674 (1991).
 71. GREENBERG, S. A., How citation distortions create unfounded authority : analysis of a citation network, *Bmj* **339** (2009).
 72. IDEKER, T. & SHARAN, R., Protein networks in disease, *Genome research* **18**, 644-652 (2008).
 73. SO, M. K., TIWARI, A., CHU, A. M., TSANG, J. T. & CHAN, J. N., Visualizing COVID-19 pandemic risk through network connectedness, *International Journal of Infectious Diseases* **96**, 558-561 (2020).
 74. OLESEN, J. M., BASCOMPTE, J., ELBERLING, H. & JORDANO, P., Temporal dynamics in a pollination network, *Ecology* **89**, 1573-1582 (2008).
 75. ZACHARY, W. W., An information flow model for conflict and fission in small groups, *Journal of anthropological research* **33**, 452-473 (1977).
 76. EULER, L., The seven bridges of Königsberg, *The world of mathematics* **1**, 573-580 (1956).
 77. NEWMAN, M. E., The structure and function of complex networks, *SIAM review* **45**, 167-256 (2003).
 78. SAMPSON, S. F., *A novitiate in a period of change : An experimental and case study of social relationships* (Cornell University, 1968).
 79. WHITE, H. C., BOORMAN, S. A. & BREIGER, R. L., Social structure from multiple networks. I. Blockmodels of roles and positions, *American journal of sociology* **81**, 730-780 (1976).
 80. HOLLAND, P. W., LASKEY, K. B. & LEINHARDT, S., Stochastic blockmodels : First steps, *Social networks* **5**, 109-137 (1983).
 81. UMEYAMA, S., An eigendecomposition approach to weighted graph matching problems, *IEEE transactions on pattern analysis and machine intelligence* **10**, 695-703 (1988).
 82. GALLO, G., LONGO, G., PALLOTTINO, S. & NGUYEN, S., Directed hypergraphs and applications, *Discrete applied mathematics* **42**, 177-201 (1993).
 83. BUTTS, C. T., Revisiting the foundations of network analysis, *science* **325**, 414-416 (2009).

-
84. DURRETT, R., *Random graph dynamics* **7** (Citeseer, 2007).
 85. ERDŐS, P., RÉNYI, A. *et al.*, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17-60 (1960).
 86. GILBERT, E. N., Random graphs, *The Annals of Mathematical Statistics* **30**, 1141-1144 (1959).
 87. FIENBERG, S. E., A brief history of statistical models for network analysis and open challenges, *Journal of Computational and Graphical Statistics* **21**, 825-839 (2012).
 88. NEWMAN, M. E., STROGATZ, S. H. & WATTS, D. J., Random graphs with arbitrary degree distributions and their applications, *Physical review E* **64**, 026118 (2001).
 89. KARINTHY, F., Chain-links, *Everything is different*, 21-26 (1929).
 90. MILGRAM, S., The small world problem, *Psychology today* **2**, 60-67 (1967).
 91. WATTS, D. J. & STROGATZ, S. H., Collective dynamics of ‘small-world’ networks, *nature* **393**, 440-442 (1998).
 92. HUBAUT, X. L., Strongly regular graphs, *Discrete Mathematics* **13**, 357-381 (1975).
 93. BARABÁSI, A.-L. & ALBERT, R., Emergence of scaling in random networks, *science* **286**, 509-512 (1999).
 94. ZUEV, K., BOGUNÁ, M., BIANCONI, G. & KRIOUKOV, D., Emergence of soft communities from geometric preferential attachment, *Scientific reports* **5**, 9421 (2015).
 95. ORGANIZATION, I. H., *Limits of oceans and seas* **23** (International Hydrographic Organization, 1953).
 96. JOHANSSON, L., JALKANEN, J.-P., KALLI, J. & KUKKONEN, J., The evolution of shipping emissions and the costs of regulation changes in the northern EU area, *Atmospheric Chemistry and Physics* **13**, 11375-11389 (2013).
 97. DUCRUET, C. & NOTTEBOOM, T., The worldwide maritime network of container shipping : spatial structure and regional dynamics, *Global networks* **12**, 395-423 (2012).
 98. GIRARDIN, R., *Ecosystem and fishers’ behaviour modelling : two crucial and interacting approaches to support Ecosystem Based Fisheries Management in the Eastern English Channel* thèse de doct. (Lille 1, 2015).
 99. IFREMER, S. d. H., Façade Mer du Nord - Manche. 2021. Activité des navires de pêche. (2021).

-
100. LEBLOND, E. *et al.*, Synthèse des flottilles de pêche 2010. Flotte de Mer du Nord-Manche-Atlantique. Flotte de Méditerranée (2012).
 101. BERTHOU, P. *et al.*, Typologies des flottes de pêche : Méthodes Ifremer-SIH (2003).
 102. GERRIOR, P., WILLIAMS, A. S. & CHRISTENSEN, D. J., Observations of the 1992 US pelagic pair trawl fishery in the Northwest Atlantic, *Marine Fisheries Review* **56**, 24 (1994).
 103. MAHÉ, K., DELPECH, J.-P., COPPIN, F. & CAPENTIER, A., Les débarquements internationaux en Manche orientale de 2000 a2004, *Direction des ressources vivante. Département Ressources Halieutiques* (2006).
 104. KENNELLY, S. J. & BROADHURST, M. K., A review of bycatch reduction in demersal fish trawls, *Reviews in Fish Biology and Fisheries* **31**, 289-318 (2021).
 105. JOO, R., ETIENNE, M.-P., BEZ, N. & MAHÉVAS, S., Metrics for describing dyadic movement : a review, *Movement ecology* **6**, 1-17 (2018).
 106. BERTRAND, M. R., DENICOLA, A. J., BEISSINGER, S. R. & SWIHART, R. K., Effects of parturition on home ranges and social affiliations of female white-tailed deer, *The Journal of wildlife management*, 899-909 (1996).
 107. LONG, J. A. & NELSON, T. A., Measuring dynamic interaction in movement data, *Transactions in GIS* **17**, 62-77 (2013).
 108. MASTROIANNI, G. & MILOVANOVIĆ, G. V., Interpolation processes : Basic theory and applications (2008).
 109. BIERNACKI, C., CELEUX, G., GOVAERT, G. & LANGROGNET, F., Model-based cluster and discriminant analysis with the MIXMOD software, *Computational Statistics & Data Analysis* **51**, 587-600 (2006).
 110. DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B., Maximum likelihood from incomplete data via the EM algorithm, *Journal of the royal statistical society : series B (methodological)* **39**, 1-22 (1977).
 111. ABDI, H. & WILLIAMS, L. J., Principal component analysis, *Wiley interdisciplinary reviews : computational statistics* **2**, 433-459 (2010).
 112. TANTARDINI, M., IEVA, F., TAJOLI, L. & PICCARDI, C., Comparing methods for comparing networks, *Scientific reports* **9**, 1-19 (2019).

-
113. SOUNDARAJAN, S., ELIASSI-RAD, T. & GALLAGHER, B., *A Guide to Selecting a Network Similarity Method* in *Proceedings of the 2014 SIAM International Conference on Data Mining* (2014), 1037-1045.
 114. MCKAY, B. D. *et al.*, Practical graph isomorphism (1981).
 115. MCKAY, B. D. & PIPERNO, A., Practical graph isomorphism, II, *Journal of symbolic computation* **60**, 94-112 (2014).
 116. READ, R. C. & CORNEIL, D. G., The graph isomorphism disease, *Journal of graph theory* **1**, 339-363 (1977).
 117. SANFELIU, A. & FU, K.-S., A distance measure between attributed relational graphs for pattern recognition, *IEEE transactions on systems, man, and cybernetics*, 353-362 (1983).
 118. NEUHAUS, M. & BUNKE, H., Self-organizing maps for learning the edit costs in graph matching, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **35**, 503-514 (2005).
 119. RISTAD, E. S. & YIANILOS, P. N., Learning string-edit distance, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, 522-532 (1998).
 120. RIESEN, K., Structural pattern recognition with graph edit distance, *Advances in computer vision and pattern recognition* (2015).
 121. GAO, X., XIAO, B., TAO, D. & LI, X., A survey of graph edit distance, *Pattern Analysis and applications* **13**, 113-129 (2010).
 122. MARZAL, A. & VIDAL, E., Computation of normalized edit distance and applications, *IEEE transactions on pattern analysis and machine intelligence* **15**, 926-932 (1993).
 123. BERLINGERIO, M., KOUTRA, D., ELIASSI-RAD, T. & FALOUTSOS, C., Netsimile : A scalable approach to size-independent network similarity, *arXiv preprint arXiv :1209.2684* (2012).
 124. LANCE, G. N. & WILLIAMS, W. T., Computer programs for hierarchical polythetic classification ("similarity analyses"), *The Computer Journal* **9**, 60-64 (1966).
 125. WILSON, R. C. & ZHU, P., A study of graph spectra for comparing graphs and trees, *Pattern Recognition* **41**, 2833-2841 (2008).

-
126. STIGLER, S. M., *The history of statistics : The measurement of uncertainty before 1900* (Harvard University Press, 1986).
127. HAEMERS, W. H. & SPENCE, E., Enumeration of cospectral graphs, *European Journal of Combinatorics* **25**, 199-211 (2004).
128. PRŽULJ, N., CORNEIL, D. G. & JURISICA, I., Modeling interactome : scale-free or geometric ?, *Bioinformatics* **20**, 3508-3515 (2004).
129. PRŽULJ, N., Biological network comparison using graphlet degree distribution, *Bioinformatics* **23**, e177-e183 (2007).
130. MILENKOVIĆ, T. & PRŽULJ, N., Uncovering biological network function via graphlet degree signatures, *Cancer informatics* **6**, CIN-S680 (2008).
131. HAYES, W., SUN, K. & PRŽULJ, N., Graphlet-based measures are suitable for biological network comparison, *Bioinformatics* **29**, 483-491 (2013).
132. YAVEROĞLU, Ö. N. *et al.*, Revealing the hidden language of complex networks, *Scientific reports* **4**, 4547 (2014).
133. SPEARMAN, C., The Proof and Measurement of Association between Two Things, *The American Journal of Psychology* **100**, 441-471, ISSN : 00029556, <http://www.jstor.org/stable/1422689> (1987).
134. WILLS, P. & MEYER, F. G., Metrics for graph comparison : A practitioner's guide, *PLOS ONE* **15**, e0228728 (2020).
135. ERDÖS, P. & RÉNYI, A., On Random Graphs I, *Publicationes Mathematicae Debrecen* **6**, 290-297 (1959).
136. CALDARELLI, G., CAPOCCI, A., DE LOS RIOS, P. & MUNOZ, M. A., Scale-free networks from varying vertex intrinsic fitness, *Physical review letters* **89**, 258702 (2002).
137. PENROSE, M., *Random geometric graphs* (OUP Oxford, 2003).
138. DAVIS, J. & GOADRICH, M., *The relationship between Precision-Recall and ROC curves in Proceedings of the 23rd international conference on Machine learning - ICML '06* (2006), 233-240.
139. HE, D., JIN, D., CHEN, Z. & ZHANG, W., Identification of hybrid node and link communities in complex networks, *Scientific reports* **5**, 1-14 (2015).

-
140. DUNNE, J. A., WILLIAMS, R. J. & MARTINEZ, N. D., Food-web structure and network theory : the role of connectance and size, *Proceedings of the National Academy of Sciences* **99**, 12917-12922 (2002).
 141. HUNTER, D. R., KRIVITSKY, P. N. & SCHWEINBERGER, M., Computational statistical methods for social network models, *Journal of Computational and Graphical Statistics* **21**, 856-882 (2012).
 142. VAN MIEGHEM, P., WANG, H., GE, X., TANG, S. & KUIPERS, F. A., Influence of assortativity and degree-preserving rewiring on the spectra of networks, *The European Physical Journal B* **76**, 643-652 (2010).
 143. TSIOTAS, D., Detecting differences in the topology of scale-free networks grown under time-dynamic topological fitness, *Scientific reports* **10**, 1-16 (2020).
 144. BARABÁSI, A.-L., ALBERT, R. & JEONG, H., Mean-field theory for scale-free random networks, *Physica A : Statistical Mechanics and its Applications* **272**, 173-187 (1999).
 145. PONCELA, J., GÓMEZ-GARDEÑES, J., FLORÍA, L. M., SÁNCHEZ, A. & MORENO, Y., Complex Cooperative Networks from Evolutionary Preferential Attachment, *PLoS ONE* **3**, e2449 (2008).
 146. BROIDO, A. D. & CLAUSET, A., Scale-free networks are rare, *Nature communications* **10**, 1-10 (2019).
 147. BELL, M. *et al.*, Network growth models : A behavioural basis for attachment proportional to fitness, *Scientific reports* **7**, 1-11 (2017).
 148. GUARE, J., in *The Contemporary Monologue Men* 89-93 (Routledge, 2016).
 149. BARRAT, A. & WEIGT, M., On the properties of small-world network models, *The European Physical Journal B-Condensed Matter and Complex Systems* **13**, 547-560 (2000).
 150. MENEZES, M. B., KIM, S. & HUANG, R., Constructing a Watts-Strogatz network from a small-world network with symmetric degree distribution, *PloS one* **12**, e0179120 (2017).
 151. BRINGMANN, K. & FRIEDRICH, T., *Exact and efficient generation of geometric random variates and random graphs* in *International Colloquium on Automata, Languages, and Programming* (2013), 267-278.

-
152. DALL, J. & CHRISTENSEN, M., Random geometric graphs, *Physical review E* **66**, 016121 (2002).
 153. R CORE TEAM, *R : A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2022), <https://www.R-project.org/>.
 154. CSARDI, G. & NEPUSZ, T., The igraph software package for complex network research, *InterJournal Complex Systems*, 1695, <https://igraph.org> (2006).
 155. DAVISON, A. C. & HINKLEY, D. V., *Bootstrap methods and their application* **1** (Cambridge university press, 1997).
 156. KRIVELEVICH, M. & SUDAKOV, B., The phase transition in random graphs : A simple proof, *Random Structures & Algorithms* **43**, 131-138 (2013).
 157. DE WINTER, J. C., GOSLING, S. D. & POTTER, J., Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes : A tutorial using simulations and empirical data. *Psychological methods* **21**, 273 (2016).
 158. GIBBONS, A., *Algorithmic graph theory* (Cambridge university press, 1985).
 159. ZURELL, D. *et al.*, The virtual ecologist approach : simulating data and observers, *Oikos* **119**, 622-635 (2010).
 160. CODLING, E. A., PLANK, M. J. & BENHAMOU, S., Random walk models in biology, *Journal of the Royal society interface* **5**, 813-834 (2008).
 161. BERGMAN, C. M., SCHAEFER, J. A. & LUTTICH, S., Caribou movement as a correlated random walk, *Oecologia* **123**, 364-374 (2000).
 162. BEDDINGTON, J., FREE, C. & LAWTON, J., Dynamic complexity in predator-prey models framed in difference equations, *Nature* **255**, 58-60 (1975).
 163. LECUN, Y., BENGIO, Y. & HINTON, G., Deep learning, *nature* **521**, 436-444 (2015).
 164. GOODFELLOW, I. *et al.*, Generative adversarial networks, *Communications of the ACM* **63**, 139-144 (2020).
 165. RUBINSTEIN, A., KUHN, H. W., MORGENSTERN, O. & VON NEUMANN, J., *Theory of Games and Economic Behavior : 60th Anniversary Commemorative Edition* (Princeton university press, 2007).

-
166. ROY, A., FABLET, R. & BERTRAND, S. L., Using generative adversarial networks (GAN) to simulate central-place foraging trajectories, *Methods in Ecology and Evolution* **13**, 1275-1287 (2022).
 167. SIMONITE, T., Software That Matches Faces Almost as Well as You Do, *Technology Review* **117**, 19-19 (2014).
 168. SILVER, D. *et al.*, Mastering the game of Go with deep neural networks and tree search, *nature* **529**, 484-489 (2016).
 169. SAGAN, C. & DRUYAN, A., *Pale blue dot : A vision of the human future in space* (Ballantine books, 2011).

Titre : Analyse du comportement collectif à l'aide de graphes, application à l'évaluation de l'indépendance des trajectoires de navires de pêche.

Mot clés : Analyse de trajectoires, Comportement collectif, Relations dyadiques, Graphes aléatoires, Comparaison de graphes, Réseaux antagonistes génératifs

Résumé : Les modèles de graphes sont couramment utilisés pour représenter les relations entre un ensemble d'entités, où l'analyse de leurs propriétés topologiques peut aider à la compréhension des phénomènes qu'ils traduisent. Lorsqu'il est question d'étudier le comportement d'un groupe de navires de pêche, les modèles de graphes peuvent être utilisés afin de mettre en perspective l'ensemble des relations synergiques ou antagonistes s'établissant entre les différents navires. Dans ce travail nous nous sommes intéressés aux propriétés d'organisations des relations deux à deux, identifiées au travers de méthodes d'analyses de trajectoires et mises en perspective à l'échelle de deux flottilles

de chalutiers. A l'aide de méthodes de comparaison de graphes et de modèles de simulation de trajectoires nous avons exploré deux approches visant à questionner la validité d'une hypothèse couramment rencontrée en sciences halieutiques, stipulant l'indépendance des navires de pêche. Nous montrons ainsi que l'hypothèse d'indépendance des trajectoires des navires de pêches ne peut pas être validée ou bien peut même être rejetée. Ces résultats pourraient intervenir dans de nombreuses approches statistiques dont les résultats servent de points d'appui au développement de stratégies de gestion des ressources halieutiques.

Title: Analysis of collective behaviour using graphs, application to the evaluation of the independence of fishing vessel trajectories.

Keywords: Trajectory analysis, Collective behaviour, Dyadic relationships, Random graphs, Graph comparison, Generative adversarial networks

Abstract: Graph models are commonly used to represent the relationships between a set of entities, where the analysis of their topological properties can help to understand the phenomena they reflect. When studying the behaviour of a group of fishing vessels, graph models can be used to put into perspective the set of synergistic or antagonistic relationships between the different vessels. In this work, we are interested in the organisational properties of pairwise relationships, identified through trajectory analysis methods and put into perspective at the scale of two trawler

fleets. Using graph comparison methods and trajectory simulation models, we explored two approaches aimed at questioning the validity of a hypothesis commonly encountered in fisheries science, stipulating the independence of fishing vessels. We show that the hypothesis of independence of fishing vessel trajectories cannot be validated or can even be rejected. These results could be used in many statistical approaches, the results of which serve as a basis for the development of fisheries management strategies.

