

Université Mohammed V-Agdal
Faculté des Sciences
de Rabat



N° d'ordre : 2690

THÈSE DE DOCTORAT

Pour l'obtention du grade de

DOCTEUR

Présentée par

Mohamed ESSALIH

**L'étude des indices topologiques, leurs applications en
QSAR/QSPR et leurs corrélations aux représentations
moléculaires «Plerograph» et «Kenograph»**

Discipline : **Sciences de l'Ingénieur**

Spécialité : **Informatique et Télécommunications**

Soutenue le 21 Décembre 2013

Devant la Commission d'Examen

Président :

M. Driss ABOUTAJDINE PES, Faculté des Sciences de Rabat

Examineurs :

M. Ahmed HAMMOUCH PES, ENSET de Rabat

M. Mohamed EL HASSOUNI PH, Faculté des Lettre et Sciences Humaines de Rabat

M. Mohamed EL MARRAKI PES, Faculté des Sciences de Rabat

Invités :

M. Mohamed Nabil SAIDI PA, INSEA de Rabat

M. Aziz YASRI Président de la société OriBase Pharma en France

Travaux réalisés au laboratoire de Recherche en Informatique et Télécommunications, Faculté des Sciences, 4 Avenue Ibn Battouta B.P. 1014 RP, Rabat.

Remerciement

Les travaux présentés dans ce mémoire ont été effectués au **L**aboratoire de **R**echerche en **I**nformatique et **T**élécommunications (**LRIT**), à la **F**aculté des **S**ciences de **R**abat (**FSR**), sous la direction de M. Mohamed EL MARRAKI.

Je tiens à remercier mon directeur de thèse, M. Mohamed EL MARRAKI, Professeur d'enseignement supérieur à la **FSR**. Qu'il trouvent d'abord mes remerciements pour la confiance qu'il m'a toujours témoignée. L'intérêt constant qu'il a pris pour ce travail, m'a permis de ne pas m'égarer dans des problèmes flous et de ne jamais perdre de vue l'essentiel, en donnant de ses temps et en acceptant de partager ses expériences.

Je tiens également à remercier M. Driss ABOUTAJDINE, Directeur du **C**entre **N**ational pour la **R**echerche **S**cientifique et **T**echnique (**CNRST**), Professeur d'enseignement supérieur à la **FSR** et directeur du **LRIT**. Sans l'environnement de recherche qu'il a su créer, je n'aurais pas pu me lancer dans la préparation de cette thèse. J'exprime ici ma profonde gratitude à son égard et l'estime respectueuse que je lui porte. Je le remercie aussi d'avoir présidé ma soutenance.

Je tiens aussi à remercier M. Ahmed HAMMOUCH, responsable du Département des Affaires Scientifiques et Techniques au **CNRST** et Professeur à l'**E**cole **N**ormale **S**upérieure de l'**E**nseignement **T**echnique (**ENSET**), de m'avoir fait l'honneur de rapporter ma thèse.

Je tiens aussi à remercier M. Mohamed EL HASSOUNI, Professeur habilité à la **F**aculté des **L**ettres et des **S**ciences **H**umaines de **R**abat (**FLSHR**) et membre permanent de **LRIT**, de m'avoir fait l'honneur de rapporter ma thèse.

Je tiens également à remercier M. Aziz YASRI, président de la société d' OriBase Pharma en France, d'avoir accepté d'être examinateur de ma thèse.

Je tiens aussi à remercier M. Mohamed Nabil SAIDI, professeur à l'Institut National de Statistique et d'Economie Appliquée (INSEA) de Rabat et membre permanent de LRIT, d'avoir accepté d'être examinateur de ma thèse.

J'ai eu l'occasion, lors de deux projets scientifiques effectués en collaboration avec M. Ivan GUTMAN et M. Boris FURTULA, Professeurs d'enseignement supérieur à la Faculté des Sciences de l'Université de Kragujevac en Serbia, de découvrir de près la recherche en représentation moléculaire (**Kenograph** et **Plerograph**). Je remercie donc toute l'équipe de ces projets, en particulier M. Ivan GUTMAN et M. Boris FURTULA.

Merci également à M. Tawfiq BARHOOM, M. Ramzi ABED, votre gentillesse et votre générosité ont énormément compté pour moi lors d'inoubliable conférence organisée à Gaza. J'espère que nous aurons prochainement l'occasion de nous rencontrer lors de futures conférences. Merci aussi à tous les chercheurs que j'ai rencontrés lors de congrès, pour leurs remarques et leurs suggestions, qui m'ont permis d'avancer dans ce travail de recherche.

Merci à mes amis de longue date que ces trois années m'ont un peu éloignés de vous. Mais que j'espère vous retrouver, bientôt, et à mes anciens compagnons du master d'Informatique et Télécommunication. Je tiens à remercier également M. Mohammed AZEEM, Mme. Fatima Zahra TELIMSANI et Mlle Keltoum AIT LAASRI de leurs remarques de forme de ce mémoire, ainsi qu'à toute personne qui a contribué de près ou de loin de la finalité de cette thèse.

Par ailleurs, c'est avec beaucoup d'émotions que je remercie mes parents pour leurs nombreux sacrifices ainsi que pour le soutien et la confiance qu'ils m'ont toujours accordé. Il m'est difficile de traduire par les mots l'affection et la gratitude que je leur réserve. Aussi, est-ce à eux que je dédie ces heures de travail, de joies et de souffrances qu'ils ont partagés. J'espère qu'ils se réjouissent de la réussite qui couronne mes efforts.

Table des matières

Table des figures	ix
Liste des tableaux	xi
Introduction générale	1
I Définitions et propriétés	5
1 Les concepts fondamentaux	7
1 Les graphes	7
2 Modes de représentation d'un graphe	12
2.1 Matrice d'adjacence	12
2.2 Matrice de distances	14
2.3 Matrice d'incidence	15
2.4 Listes d'adjacences	15
2.5 Matrice Laplacienne	16
3 La théorie des graphes moléculaires	17
3.1 Les concepts généraux	17
3.2 Les graphes moléculaires	19
3.2.1 Définition	20
3.2.2 La représentation d'une molécule	20

2	La prédiction de propriétés et activités moléculaires	21
1	Les descripteurs moléculaires	22
1.1	Définition	22
1.2	Les types de descripteurs moléculaires	23
1.2.1	Les descripteurs 1 <i>D</i>	23
1.2.2	Les descripteurs 2 <i>D</i>	23
1.2.3	Les descripteurs 3 <i>D</i>	24
2	Les études QSAR/QSPR	25
2.1	Définition de QSAR/QSPR	26
2.2	Préparation des données d'entrée	26
2.3	Génération des descripteurs moléculaires	27
2.4	Sélection des descripteurs les plus adaptés	27
2.4.1	L'analyse en composantes principales	28
2.4.2	La méthode de régression des moindres carrés partiels	28
2.4.3	sélection de variables pertinentes	29
2.4.4	Le regroupement de plusieurs indices	29
2.5	Relation des descripteurs à la grandeur	30
2.6	la validation de modèle	30
2.7	Autres méthodes de QSAR/QSPR	31
2.7.1	Méthode de contribution de groupes	31
2.7.2	Analyse comparative de champs moléculaires (CoMFA)	33
3	Le criblage virtuel	34
3.1	Le criblage virtuel basé sur la structure de la cible	35
3.2	Le criblage virtuel basé sur la structure du ligand	35
3.2.1	Les méthodes locales	36
3.2.2	Les méthodes globales	36
II	Les méthodes de calcul de l'indice de Wiener et ses extensions	39
3	L'indice de Wiener	41
1	La définition de base et ses applications	42
1.1	L'histoire de l'indice de Wiener	42
1.2	La définition de l'indice de Wiener	43
1.3	L'utilisation de définition de l'indice de Wiener	44
1.3.1	Le graphe Star-trees	44

1.3.2	Le graphe Path-trees	45
1.4	Les utilités de l'indice de Wiener	46
2	La formule de l'indice de Wiener améliorée	47
2.1	La formule de l'indice de Wiener en fonction du diamètre	47
2.2	L'utilisation de la formule avec des graphes à faible diamètre	50
2.2.1	Les graphes de diamètre 2	50
2.2.2	Les graphes de diamètre 3	52
2.2.3	Les graphes de diamètre 4	54
2.2.4	Quelques graphes classique de diamètre supérieur à 4	56
2.3	L'utilisation de la formule avec les graphes composés	59
2.3.1	Le produit Corona de deux graphes Cycle	59
2.3.2	Le produit Corona des graphes Cycle et Path	61
2.4	L'indice de Wiener pour le graphe Cobweb	63
3	L'indice de Wiener pour les arbres	65
3.1	Les formules directes	65
3.1.1	Les points de branchement et les segments	66
3.1.2	Les valeurs propres de la matrice Laplacienne	69
3.2	Les méthodes récursives	69
3.2.1	L'arbre de Thorn	72
3.2.2	L'arbre à $k - sous$ divisions d'un arbre	75
4	Le problème de l'inverse de Wiener	76
4	Les extensions de l'indice de Wiener	77
1	L'indice de Hyper-Wiener	77
1.1	L'invariant de Wiener-Type	78
1.1.1	Définition	78
1.1.2	Quelques propriétés de l'invariant de Wiener-Type	79
1.2	L'indice de Hyper-Wiener	81
1.2.1	Définition	82
1.2.2	La formule de l'indice d'Hyper-Wiener améliorée	82
1.2.3	La formule de l'indice d'Hyper-Wiener en fonction du diamètre	83
1.2.4	L'utilisation de la formule avec des graphes à faible diamètre	84
2	L'indice de degré de distance	86
2.1	L'indice topologique moléculaire de Schultz (MTI)	87
2.1.1	Définition	87

Table des matières

2.1.2	Quelques propriétés de MTI	88
2.2	L'indice de degré de distance $DD(G)$	88
2.2.1	Définition	89
2.2.2	Quelques propriétés de $DD(G)$	91
2.2.3	La réécriture de la formule de l'indice de degré de distance	92
3	L'indice de Terminal de Wiener	98
3.1	Définition	98
3.2	Quelques propriétés	99
3.2.1	Les propriétés simples	99
3.2.2	Les limites de l'indice de Terminal de Wiener	100
3.2.3	L'arbre régulier T^* correspondant à un arbre T	102
III La corrélation entre Plerograph et Kenograph		109
5 Les représentations Kenograph et de Plerograph		111
1	La terminologie des représentations graphiques	112
1.1	Historique	112
1.2	Les représentations Kenograph et Plerograph	113
1.2.1	Définitions	113
1.2.2	Les vrais terms	113
1.2.3	La négligence du Plerograph	114
1.2.4	Le cas générale	115
2	La relation entre les types Kenograph & Plerograph	116
2.1	La corrélation entre les Walks en Plerograph et Kenograph	116
2.2	La corrélation entre les TI basés sur la distance	118
2.2.1	Le cas de l'invariant de Wiener-Type	118
2.2.2	Le théorème 5.1 pour l'indice de Wiener	120
2.2.3	La relation entre $W(Pl)$, $W(Ke)$, $TW(Pl)$ et $TW(Ke)$	123
2.3	La corrélation entre les TI basés sur le degré	126
2.3.1	La formule de base	126
2.3.2	Le cas de Plerograph	127
2.3.3	L'étude de quelques TI basé sur les valeurs propres	127
3	La dépendance structurelle de $TW(G)$	129
3.1	Problématique & discussion	129
3.1.1	Problématique	129

3.1.2	Discussion	130
3.2	La dependance de l'indice de Terminal de Wiener au Kenograph	134
3.2.1	Le résultat principal	134
3.2.2	Discussion	135
3.3	La dependance de l'indice de Terminal de Wiener au Plerograph	136
3.3.1	Le résultat principal	136
3.3.2	Discussion	137
 Conclusion générale et perspectives		 139
 Qu'est-ce que la classification ?		 143
1	Quelques exemples introductifs	143
2	Le système de reconnaissance de forme	144
2.1	Module d'acquisition	144
2.2	Module de prétraitement	145
2.3	Modules complémentaires	145
2.4	Module de classification	145
2.5	Exemple d'un système de reconnaissance de formes	145
3	Le problème de classification mathématiquement	146
4	Quelques exemples distinctifs	146
4.1	La classification exacte	147
4.1.1	Le rangement de pièces de monnaie	147
4.1.2	La notation des restaurants dans les guides touristiques	148
4.2	La classification probabiliste	149
4.2.1	La discrimination des femmes des hommes	149
5	Résolution des problèmes de classification	149
5.1	Le principe de résolution	150
5.2	La procédure de résolution par apprentissage	150
5.2.1	Première étape	150
5.2.2	Deuxième étape	150
5.2.3	Troisième étape	151
5.2.4	Dernière étape	151
5.3	Les précautions à prendre en compte lors de l'apprentissage	151
6	Qu'est-ce que la classification ?	153
6.1	Système expert	153

Table des matières

6.2	Système d'apprentissage	154
6.2.1	Les méthodes statistiques	154
6.2.2	Les méthodes non statistiques	154
6.2.3	Les applications de l'apprentissage	155
	Références bibliographiques	157

Table des figures

1.1	Exemple de graphes non orientés, connexes, planaires et finis	8
1.2	Exemple de graphes orientés et non orientés finis	9
1.3	Exemple d'un arbre	16
1.4	les composés chimiques de 2, 2, 4, 6-tetramethylheptane (I) et diphenylene (II)	19
1.5	Les représentations Plerograph (Pl) et kenograph (Ke) de 2, 2, 3, 5-tetramethylhexane	20
2.1	Exemple de deux molécules, isomères de constitution	32
3.1	Exemple de graphe Star-Trees	43
3.2	Exemple de graphe Path-Trees	44
3.3	Exemple du graphe Maximale planaire \mathcal{E}_n d'ordre n	48
3.4	Exemple de graphe Fan \mathcal{F}_n d'ordre n	50
3.5	Exemple du graphe Wheel \mathcal{W}_n d'ordre n	51
3.6	Exemple du graphe Butterfly \mathcal{B}_n d'ordre n	51
3.7	Exemple du graphe Crystal \mathcal{C}_n d'ordre n	52
3.8	Exemple de graphe Double-Crystal \mathcal{DC}_n d'ordre n	53
3.9	Exemple de graphe Sunflower \mathcal{FL}_n d'ordre n	54
3.10	Exemple de graphe Double-Wheel \mathcal{DW}_n d'ordre n	55
3.11	Exemple de graphe Cycle \mathcal{O}_n d'ordre n	57
3.12	Exemple de graphe Grid \mathcal{G}_n d'ordre n	58
3.13	Exemple de graphe Path \mathcal{P}_n d'ordre n	58
3.14	Exemple du produit Corona de deux graphes Cycle $C_{n_1} \circ C_{n_2}$	60
3.15	Exemple du produit Corona des graphe Cycle et Path $C_{n_1} \circ P_{n_2}$	62

Table des figures

3.16	Exemple du graphe Cobweb WB_n d'ordre n formé de 3 niveaux	64
3.17	Un arbre avec son point de branchement v et trois sous-arbre T_i	67
3.18	Les arbres T_1 et T_2 , connexes par un k sommets	70
3.19	Exemple d'identification des arbres	71
3.20	Exemple du graphe fasciagraph F	72
3.21	Exemple d'un arbre et le graphe Thorn correspond	73
4.1	Exemple de graphe Star S_n d'ordre n	79
4.2	L'arête créant les sous-graphes F_i de même nombre de sommets et différentes $DD(F_i)$	89
4.3	Le graphe montrant la non-monotonie de $DD(G)$ vis-à-vis l'ajout d'une arête	90
4.4	Exemple du graphe Star-Graph $G_{m,n}$ d'ordre n	97
4.5	Les arbres Starlike de degré 4 et d'ordre 12	101
4.6	L'arbre T et l'arbre régulier T^* correspondant à T pour $k = \Delta = 6$	103
5.1	La rivière Pregel et l'île de Kneiphof	113
5.2	Exemple de graphe Thorn pour $\alpha_{max} = 3$ pour 2 et $\alpha_{max} = 4$ pour 1	115
5.3	La corrélation entre $w(Ke)$ et $w(Pl)$ pour les isomères d'Octane	117
5.4	La corrélation entre les indices de Harary du Pl et Ke	121
5.5	La corrélation entre les indices de l'inverse de Wiener de Pl et Ke	121
5.6	Les indices de Wiener de Ke et Pl pour les isomères de l'Octane	122
5.7	$W(Ke)$ et $W(Pl)$ pour les isomères de Monocyclooctanes	122
5.8	Exemple de nouvelle notation	123
5.9	La relation entre W et TW pour les arbres d'ordre 8	124
5.10	La corrélation entre $\lambda_1(Ke)$ et $\lambda_1(Pl)$ pour les isomères d'Undecanes ($C_{12}H_{26}$)	128
5.11	La corrélation entre $EE(Ke)$ et $EE(Pl)$ pour les isomères d'Undecanes ($C_{12}H_{26}$)	128
5.12	La corrélation entre $TW(Ke)$ et $TW(Pl)$ pour les isomères de Nonanes	130
5.13	2 Kenographs de $n = 15$ avec des p différentes et de même $TW(Ke)$	133
5.14	2 Kenographs de $n = 14$ avec des $p_1 \leq p_2$ mais $TW(Ke_{p_1}) \geq TW(Ke_{p_2})$	133
5.15	Le schéma explicatif de transformation Ke à Ke'	135
5.16	Exemple du Kenograph $Ke(a_1, a_2, a_3)$ et son Plerograph	137
A.1	chaîne d'un dispositif de classification	146
A.2	Un problème de classification	152

Liste des tableaux

1.1	Les listes d'adjacences du graphe G_2 de la Figure 1.2	15
2.1	Décomposition en groupes de deux isomères de constitution	32
5.1	La corrélation entre Les Walks du Ke et du Pl pour les octanes	117
5.2	Certaines valeurs de a et b de formule 5.2 en fonction de n et λ	119
5.3	Les valeurs de λ et les indices correspondants	120
5.4	les Coefficients de corrélation de formule 5.2 pour quelques λ	120

Résumé

Les processus qui mènent à la découverte de nouveaux médicaments sont longs et fastidieux, et les taux de succès sont relativement faibles. L'identification de candidats par le biais de tests expérimentaux s'avère coûteuse, et nécessite de connaître en profondeur les mécanismes d'action de la protéine visée afin de mettre en place des essais efficaces. La modélisation de propriétés et d'activités de molécules peut considérablement accélérer ces processus en permettant une évaluation rapide de plusieurs milliers de molécules afin de déterminer lesquelles sont les plus susceptibles de se lier à une activité biologique ou propriété physique/chimique d'une molécule. L'un des challenges de la chémo-Informatique est d'être capable de décrire de manière simple des molécules afin de pouvoir les utiliser dans des études de similarité ou de pouvoir prédire leur activité en se basant sur les informations contenues dans les composés déjà connus. De nombreuses recherches ont été menées, au cours des dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules, et ces structures elles-mêmes, en un ensemble de nombres réels appelés indices topologiques ; une fois que ces nombres sont disponibles, il est possible d'établir une relation entre ceux-ci et une propriété ou activité moléculaire, à l'aide d'outils de modélisation classiques. Ces indices topologiques réalisent de ce fait un codage de l'information chimique en un vecteur de réels.

Le premier objectif de ce travail était d'améliorer la façon de calculer quelques principaux indices topologiques tels que : les indices de Wiener, de degré de distance, d'Hyper-Wiener et de Terminal de Wiener, en proposant des formules utilisant juste le nombre de paires de sommets ayant la même distance k ($d_G(k)$) pour les calculer. On a utilisé ces formules avec des graphes de littérature, de différents diamètres, ou que nous les avons conçus afin de les vérifier. L'étude de ces indices nous a fourni une vision globale sur leurs définitions et nous l'a exploité afin de regrouper ces indices en des groupes selon la matière première de leurs définitions (distance entre deux sommets, degrés des sommets, ...) et donc, donné pour chaque groupe d'indices une relation qui lui convient. Ce regroupement améliorera, sans doute, les techniques de méthodes de la sélection des variables qui font parties des méthodes de modélisation. Puis, on a étudié la notion de la représentation moléculaire (**Kenograph** et **Plerograph**) car elle se considère dans la littérature comme le premier pas de calcul des indices topologiques en donnant d'autres raisons justifiant la domination de type Kenograph depuis la création de la théorie des graphes moléculaire.

Mots clés : Descripteurs moléculaires, Indices topologiques, Graphe, Kenograph, Plerograph, Indice de Wiener.

Abstract

The process of drug discovery is long and tedious. Besides, it is relatively inefficient in terms of hit rate. The identification of candidates through experimental testing is expensive and requires extensive data on the mechanisms of the target protein in order to develop efficient assays. The modeling of the molecular properties and activities can considerably accelerate the process by quickly evaluating large databases of compounds and determining the most likely to bind to a biological activity or physical/chemical property of a molecule. One of the challenges faced by Chemo-informatics is the following : to be able to describe compounds in a simple way, in order to use them in similarity studies or to predict their activity, based on information contained in already known compounds. In recent decades, much researches have been carried out to find the best way of representing the information contained in the structure of molecules, and these structures themselves, into a set of real numbers called topological indices ; Once these numbers are available, it is possible to establish a relationship there between and a molecular property or activity, using the traditional modeling tools. These topological indices realize this is an encoding of chemical information into a vector of real numbers.

The first objective of this work were to improve the manner of calculating some main topological indices such as : the Wiener index, the degree distance index, the Hyper-Wiener index and the Terminal Wiener index, by proposing formulas using just the number of pairs of vertices having the same distance k ($d_G(k)$) to calculate them. We applied these formulas on graphs of literature, having different diameters, or that we have designed in order to verify it. The study of these indices has given us a global vision about their definitions and we exploited it to bring together these indices into groups according to the raw material of their definitions (distance between two vertices, vertex degrees, ...) and therefore, given for each group of these indices a relationship that appropriate it. This grouping will improve undoubtedly the technical methods of the selection of variables which that are part of the modeling methods. The second objective is to study the concept of the molecular representation (**Kenograph** and **Plerograph**) because it is considered in the literature as the first step to calculate the topological indices, and giving an other reasons for the dominance of the **Kenograph** type since the creation of the of molecular graphs.

Keywords : Moléculars descriptors, Topological indices, Graph, Kenograph, Plerograph, Wiener index

Introduction générale

La Bio-informatiques et la Chémo-informatique sont des disciplines devenues indispensables à la découverte de médicaments. De nos jours, les industries pharmaceutiques consacrent près de 10% de leur budget de recherche et développement, à la recherche de médicaments assisté par ordinateur [127]. Cette émergence peut s'expliquer à la fois par le développement des architectures et techniques de calculs mais aussi par le faible coup qu'engendrent des analyses sur ordinateur par rapport à des tests dans les laboratoires biologiques. Il est admis que, en moyenne, pour une molécule qui arrive sur le marché en tant que médicament innovant, 10000 molécules sont synthétisées et testées. De plus, le développement d'un médicament demande généralement entre 10 et 15 ans de recherches. Il s'agit en effet de trouver une molécule qui doit à la fois présenter des propriétés thérapeutiques particulières, et posséder le minimum d'effets secondaires indésirables. Le prix de revient d'un médicament est essentiellement dû à ces synthèses longues, coûteuses et finalement inutiles [16].

Pour toutes ces raisons, les industries pharmaceutiques s'orientent vers de nouvelles méthodes de recherche, qui consistent à prédire les propriétés et activités de molécules avant même que celles-ci ne soient synthétisées. Deux disciplines de la «chimie computationnelle» se sont développées en réponse à ce besoin : les relations structure-activité ou *QSAR* (Quantitative Structure-Activity Relationships), et les relations structurepropriété ou *QSPR* (Quantitative Structure-Property Relationships). Elles consistent essentiellement en la recherche de similitudes entre molécules dans de grandes bases de données de molécules existantes dont les propriétés sont connues. La découverte d'une telle relation permet de prédire les propriétés physiques et chimiques et l'activité biologique de composés, de développer de nouvelles théories ou de comprendre les phénomènes observés. Elle permet également de guider les synthèses de nouvelles molécules, sans avoir à les réaliser, ou à analyser des familles entières de composés.

Les relations entre les structures des molécules et leurs propriétés ou activités sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. De nombreuses recherches ont été menées, au cours des dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules, et ces structures elles-mêmes, en un ensemble de nombres réels appelés indices topologiques ; une fois que ces nombres sont disponibles, il est possible d'établir une relation entre ceux-ci et une propriété ou activité moléculaire, à l'aide d'outils de modélisation classiques. Ces descripteurs numériques réalisent de ce fait un codage de l'information chimique en un vecteur de réels. On en dénombre aujourd'hui plus de 3000 familles des indices topologiques, qui quantifient des caractéristiques physico-chimiques ou structurelles de molécules. Ils peuvent être obtenus de manière empirique ou non-empirique, mais les descripteurs calculés, et non mesurés, sont à privilégier : ils permettent en effet d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs de la modélisation. Il existe cependant quelques descripteurs mesurés : il s'agit généralement de données expérimentales plus faciles à mesurer que la propriété ou l'activité à prédire (coefficient de partage eau-octanol [36], polarisabilité, ou potentiel d'ionisation). Avant donc toute modélisation, il est nécessaire de calculer ou de mesurer un grand nombre de descripteurs différents, car les mécanismes qui déterminent l'activité d'une molécule ou une de ses propriétés sont fréquemment mal connus. Il faut ensuite sélectionner parmi ces variables celles qui sont les plus pertinentes pour la modélisation.

Ce rapport est découpé en trois parties :

- La première partie est constituée de deux chapitres et consacrée à l'étude de la théorie des graphes et ses applications aux différentes méthodes de *QSAR/QSPR*. On commence dans le chapitre 1 par introduire la notion de graphe et donner quelques définitions et propriétés relatives aux graphes. On s'intéressera plus particulièrement aux graphes moléculaires. Le chapitre 2 traite lui de la notion des descripteurs moléculaires ainsi que du rôle important que va jouer ces derniers dans les études *QSAR/QSPR*.
- La deuxième partie explique en détaille les différentes méthodes de calcul d'une catégorie particulière des descripteurs moléculaires (les indices topologiques, en particulier l'indice de Wiener et ses extensions). Cette partie comporte l'essentiel de nos contributions [77][1][150][151][3] [5][4][2][6][7]. Elle est découpée en deux chapitres, le premier chapitre (chapitre 3) est dédié au calcul de l'indice de Wiener en utilisant plusieurs méthodes (utilisation directe de la définition, utilisation du théorème basé sur le diamètre). On étudiera plus particulièrement les arbres (graphe connexe et sans cycle). Le deuxième chapitre de cette partie (c'est-à-dire le chapitre 4) est consacré à l'étude de quelques autres indices topologiques qui

sont en relation avec l'indice de Wiener (comme les indices de Hyper-Wiener, de degré de distance et Terminal de Wiener).

- La dernière partie de ce rapport (partie *III*) est formée d'un seul chapitre (chapitre 5). Elle est consacrée à l'étude des représentations Kenograph *Ke* et Plerograph *Pl*. Après avoir donné les définitions des ces représentations, on a étudié la relation entre eux. Pour pouvoir trouver cette corrélation il nous a fallu utiliser l'indice de Terminal de Wiener. Les travaux de ce chapitre ont fait l'objet de deux contributions avec M. Ivan Gutman [117][108].

On termine ce rapport par donner une conclusion générale et quelques perspectives. Et à la fin de ce manuscrit se trouve une annexe qui relate la notion de classification et ses différentes méthodes statistique (réseaux de neuron, K-plus proche voisin, ...).

Première partie

Définitions et propriétés

Chapitre 1

Les concepts fondamentaux

Introduire une nouvelle matière n'est pas toujours chose plaisante car il s'agit souvent d'une accumulation de définitions ! Et c'est la situation rencontrée dans ce chapitre. Nous allons donc agrémenter cette présentation, autant que faire se peut, d'exemples mettant en lumière l'intérêt pratique de la théorie des graphes. Comme l'indique le titre de ce chapitre, nous nous intéressons aux concepts fondamentaux des graphes, en particulier aux leurs différentes définitions, quelques leurs propriétés. Nous allons aussi voir une petite introduction des graphes chimiques dans laquelle nous nous intéresserons à leur importance pratique à travers les descripteurs topologique, et dans chaque section nous donnerons des exemples qui vont clarifier ces définitions.

1 Les graphes

Définition 1.1. Soient V un ensemble (fini ou infini) et E une partie de $V \times V$ (i.e., une relation sur V). Le graphe $G = (V, E)$ est la donnée du couple (V, E) . Les éléments de V sont appelés les sommets ou noeuds de G , et ceux de E sont appelés les arcs ou arêtes de G . Si V est fini, on parlera de graphe fini [29, 34, 133, 159, 163, 197]. La Figure 1.1 montre trois exemples des graphes finis : G_1 , G_2 et G_3 .

Définition 1.2. L'ordre d'un graphe, noté n , est le nombre de ses sommets.

Exemple 1.1. Les ordres des graphes G_1 , G_2 et G_3 de la Figure 1.1, sont $n_1 = 12$, $n_2 = 12$ et $n_3 = 11$ respectivement.

Définition 1.3. La taille d'un graphe, notée m , est le nombre de ses arêtes.

Exemple 1.2. Les tailles des graphes G_1 , G_2 et G_3 de la Figure 1.1, sont $m_1 = 11$, $m_2 = 14$ et $m_3 = 13$ respectivement.

Chapitre 1. Les concepts fondamentaux

Définition 1.4. Le voisinage d'un sommet, noté ν , est l'ensemble de tous ses sommets adjacents. Sa taille est $|\nu|$.

Exemple 1.3. Le voisinage du sommet w_2 du graphe G_3 de la Figure 1.1 est $\nu(w_2) = \{w_1, w_3, w_7, w_8, w_9, w_{10}, w_{11}\}$, et $|\nu(w_2)| = 7$.

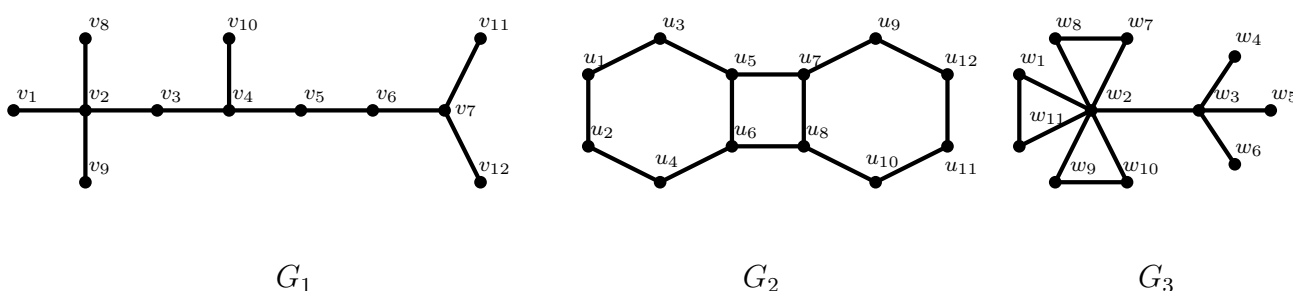


Figure 1.1 – Exemple de graphes non orientés, connexes, planaires et finis

Remarque 1.1. Vis-à-vis les deux types de graphe (orienté ou non orienté), on a les remarques suivantes :

1. Observons que l'ordre au sein des couples appartenant à E est intrinsèquement présent. On parlera donc parfois de graphe orienté ou de graphe dirigé (digraph). Si V est l'ensemble des sommets du graphe G et si $a = (v_i, v_j) \in E$ est une arête de ce graphe, on pourra alors parler de l'origine v_i et de la destination v_j de l'arc a . On dit que v_i et v_j sont les extrémités de l'arc a et que a relie v_i à v_j . Si $b = (v_i, v_i)$, on parle généralement de la boucle b .
2. Il est souvent commode de donner une représentation sagittale d'un graphe. Les sommets sont représentés par des points et si la paire de sommets (v_i, v_j) représente un arc, alors on trace une flèche de v_i vers v_j .
3. Dans un graphe orienté (resp. non orienté), nous nous efforcerons de parler d'arcs (resp. d'arêtes), pour désigner la liaison entre deux sommets dans ce graphe.

Définition 1.5. Soit $G = (V, E)$ un graphe. Si E est une relation symétrique sur V , on dira que G est un graphe non orienté. Autrement dit, G est non orienté si :

$$\forall v_i, v_j \in V : (v_i, v_j) \in E \Rightarrow (v_j, v_i) \in E \quad (1.1)$$

Remarque 1.2. Dans le cas des graphes non orientés :

1. On simplifie la représentation sagittale de G en traçant simplement un segment entre v_i et v_j .
2. Pour alléger l'écriture, on identifiera les arcs (v_i, v_j) et (v_j, v_i) avec une unique arête non orientée donnée par la paire $\{v_i, v_j\}$.

Exemple 1.4. Voici les exemples suivants :

- Les graphes qui figurent dans la Figure 1.1 sont tous des graphes non orientés, et le graphe G_1 de la Figure 1.2 est un graphe orienté.
- Le couple de sommets $a = (v_1, v_2)$ du graphe G_1 de la Figure 1.2 est un arc, qui diffère de l'arc $a' = (v_2, v_1)$.
- Le couple de sommets $b = \{v_1, v_2\}$ du graphe G_1 de la Figure 1.1 est une arête, et c'est la même que l'arête $b = \{v_2, v_1\}$.
- Le couple de sommets $c = \{w_3, w_3\}$ du graphe G_3 de la Figure 1.2 est une boucle.

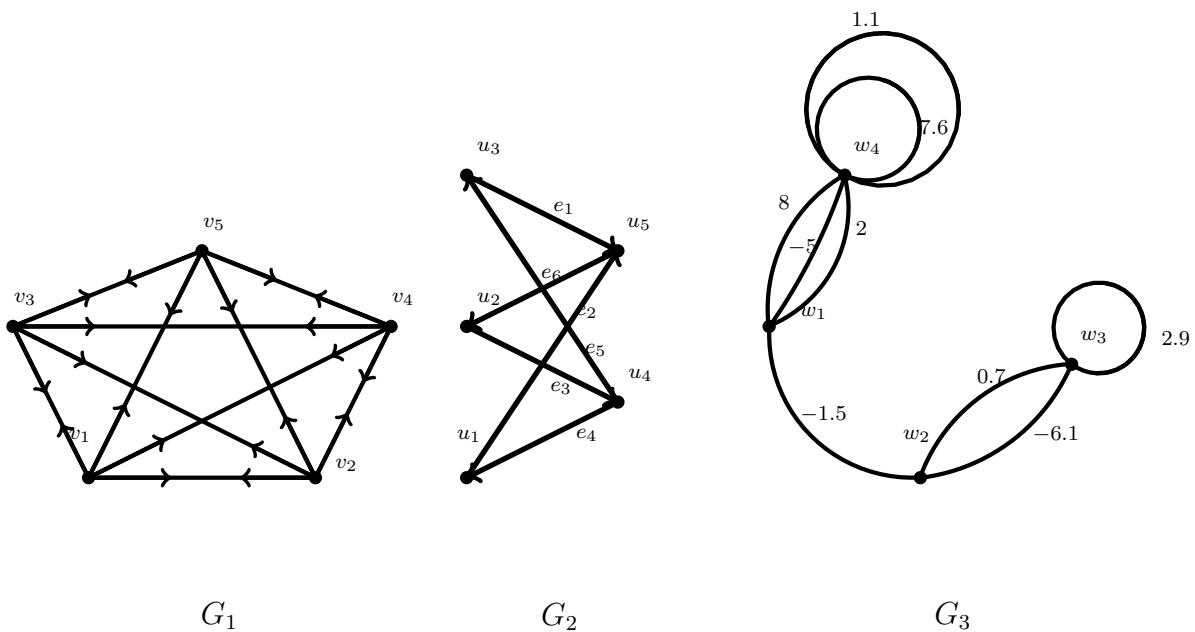


Figure 1.2 – Exemple de graphes orientés et non orientés finis

Définition 1.6. Un multi-graphe est un graphe contenant des boucles et/ou plusieurs arêtes reliant les mêmes sommets.

Définition 1.7. Un graphe $G = (V, E)$ est dit simple s'il ne s'agit pas d'un multi-graphe et s'il ne contient pas de boucle.

Chapitre 1. Les concepts fondamentaux

Définition 1.8. Un graphe planaire est un graphe qu'on peut dessiner sur une surface plate sans que ses arêtes ne se croisent. Les graphes qu'on ne peut pas dessiner sans croisement sont dits non planaires [29, 34, 133, 159, 163, 197].

Définition 1.9. Un graphe $G = (V, E)$ est dit bi-parti si V peut être partitionné en deux ensembles V_1 et V_2 de manière que $E \subseteq V_1 \times V_2$. Si $|V_1| = n_1$, $|V_2| = n_2$ et $E = V_1 \times V_2$, alors on parle du graphe bi-parti complet et il est noté K_{n_1, n_2} .

Exemple 1.5. Voici les exemples suivants :

- Tous les graphes qui figurent dans la Figure 1.1 sont des graphes simples et planaires. Les graphes G_1 et G_2 de la Figure 1.2 sont deux graphes non planaires.
- Le graphe G_3 de la Figure 1.2 est un multi-graphe.
- Le graphe G_2 de la Figure 1.2 est un graphe bi-parti.
- Le graphe G_1 de la Figure 1.1 est un graphe simple.

Définition 1.10. Deux arêtes sont dites adjacentes s'ils ont au moins une extrémité en commun.

Exemple 1.6. Les arêtes $\{v_2, v_3\}$ et $\{v_3, v_4\}$, du graphe G_1 figurant dans la Figure 1.1, sont adjacentes.

Définition 1.11. Deux sommets $v_i, v_j \in V$ sont adjacents si l'arête $\{v_i, v_j\}$ appartient à E . On dit aussi qu'ils sont voisins. L'ensemble des voisins de v se note $\nu(v)$.

Exemple 1.7. Les sommets u_7 et u_8 , du graphe G_2 figurant dans la Figure 1.1, sont adjacents.

Définition 1.12. Un graphe $G = (V, E)$ est complet, si toutes les paires de ses sommets sont adjacentes. Un graphe complet d'ordre n est noté K_n (le K est en l'honneur de Kuratowski, un pionnier de la théorie des graphes).

Définition 1.13. On appelle degré du sommet u , noté par $deg(u) = |\nu(u)|$, le nombre d'arêtes incidentes à u , on suppose en outre que les boucles apportent une double contribution au degré d'un sommet. Le degré d'un graphe est le degré maximum de tous ses sommets.

Exemple 1.8. Voici les exemples suivants :

- Les degrés des sommets v_1, u_7 et w_2 des graphes G_1, G_2 et G_3 de la Figure 1.1, sont, respectivement, de $deg_{G_1}(v_1) = 4$, $deg_{G_1}(u_7) = 3$ et $deg_{G_3}(w_2) = 7$.
- Le graphe G_1 de la Figure 1.2 est un graphe complet.

Lemme 1.1. [133] Si $G = (V, E)$ est un multi-graphe non orienté, alors :

$$\sum_{u \in V} deg(u) = 2|E| \tag{1.2}$$

Définition 1.14. Un chemin dans un graphe orienté est une suite de sommets reliés les uns aux autres par des arcs. Sa longueur est le nombre de ses arcs, ou le nombre de ses sommets moins un. Un chemin simple ne peut pas visiter le même arc plus d'une fois. Un chemin fermé a pour dernier sommet le premier.

Définition 1.15. Une chaîne dans un graphe non orienté est une suite de sommets reliés par des arêtes. Sa longueur est le nombre de ses arêtes, ou le nombre de ses sommets moins un. Une chaîne élémentaire ne peut pas visiter le même sommet deux fois. Une chaîne simple ne peut pas visiter la même arête deux fois.

Exemple 1.9. Voici les exemples suivants :

- La suite de sommets $s_1 = \{w_1, w_2, w_7, w_8, w_2, w_1\}$ du graphe G_3 de la Figure 1.1, forme une chaîne de longueur $l_1 = 5$.
- La suite de sommets $s_2 = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_{11}\}$ du graphe G_1 de la Figure 1.1, forme une chaîne élémentaire de longueur $l_2 = 7$.
- La suite de sommets $s_3 = \{u_1, u_3, u_5, u_7, u_9, u_{12}, u_{11}\}$ du graphe G_2 de la Figure 1.1 forme une chaîne simple de longueur $l_3 = 6$.

Définition 1.16. Un graphe connexe est un graphe dans lequel chaque paire de sommets est reliée par une chaîne. Un graphe qui n'est pas connexe est dit non connexe, et se décompose en composantes connexes.

Exemple 1.10. Tous les graphes de la Figure 1.1 sont des graphes connexes.

Définition 1.17. Dans un graphe orienté, un circuit est un chemin fermé et simple.

Définition 1.18. Dans un graphe non orienté, un cycle est une chaîne élémentaire, simple et dont ses extrémités coïncident.

Exemple 1.11. Voici les exemples suivants :

- La suite de sommets, $c = \{u_1, u_3, u_5, u_6, u_4, u_2\}$ du graphe G_2 de la Figure 1.1, est un cycle.
- La suite de sommets $c_1 = \{v_1, v_2, v_3\}$ du graphe G_1 de la Figure 1.2, est un circuit de longueur $l_1 = 3$.
- La suite de sommets $c_2 = \{v_1, v_2, v_3, v_4, v_5, v_1\}$ du graphe G_1 de la Figure 1.2 est un chemin fermé, et qui est aussi un circuit, de longueur $l_2 = 5$.

Définition 1.19. La distance entre deux sommets u et v , noté par $d(v, u)$, est la longueur de la plus courte chaîne entre eux.

Définition 1.20. Le diamètre d'un graphe $G = (V, E)$, noté par $D(G)$ est la plus longue distance entre deux sommets de ce graphe.

Exemple 1.12. Voici les exemples suivants :

- La distance entre les sommets v_1 et v_7 , du graphe G_1 de la Figure 1.1, est $d(v_1, v_7) = 6$.
- Les diamètres des graphes G_1 , G_2 et G_3 sont respectivement $D(G_1) = 7$, $D(G_2) = 6$ et $D(G_3) = 3$.

Définition 1.21. L'arbre est un graphe connexe et qui ne contenant aucun cycle.

Définition 1.22. La forêt est un graphe qui ne contient aucun cycle. Les composantes connexes d'une forêt sont des arbres.

Définition 1.23. La feuille est un sommet de degré un. Aussi appelé sommet pendent.

Définition 1.24. Un graphe G est une arborescence s'il existe un sommet R appelé racine de G tel que, pour tout sommet S de G , il existe un chemin et un seul de R vers S .

Exemple 1.13. Voici les exemples suivants :

- Le graphe G_1 de la Figure 1.1, est un arbre.
- Les deux graphes G_2 et G_3 ne sont pas des arbres.
- Les sommets $v_1, v_8, v_{10}, v_{12}, v_9$ et v_{11} sont des feuilles du graphe G_1 de la Figure 1.1.

2 Modes de représentation d'un graphe

L'essor de la théorie des graphes est essentiellement dû à l'avènement de puissants calculateurs. Il est donc légitime de s'intéresser à la manière de représenter des graphes au sein d'un ordinateur. Plusieurs modes de représentation peuvent être envisagés selon la nature des traitements qu'on souhaite appliquer au graphe considéré. Dans cette section nous allons traiter quelques-uns d'entre eux et pour en savoir plus, voici quelques références [141, 159, 179, 197].

2.1 Matrice d'adjacence

Définition 1.25. Considérons un graphe $G = (V, E)$ d'ordre n . Sa matrice d'adjacence notée $A = (a_{ij})$ est une matrice de dimension $n \times n$, tel que

$$a_{ij} = \begin{cases} 1, & \text{si } (i, j) \in E; \\ 0, & \text{sinon.} \end{cases}$$

Remarque 1.3. Cette matrice a plusieurs caractéristiques :

1. Elle est carrée.

2. Il n'y a que des zéros sur la diagonale. Un 1 sur la diagonale indiquerait une boucle.
3. Elle est symétrique : $a_{ij} = a_{ji}$. On peut dire que la diagonale est son axe de symétrie.
4. Une fois qu'on fixe le nombre de sommets, il existe une matrice d'adjacences unique pour chaque graphe. Celle-ci n'est la matrice d'adjacence d'aucun autre graphe.
5. Un graphe orienté a une matrice d'adjacence quelconque, alors qu'un graphe non orienté possède une matrice d'adjacence toujours symétrique.
6. Ce mode de représentation engendre des matrices creuses. Cependant la recherche de chemin ou de chaîne s'effectue aisément avec une telle représentation.
7. Le graphe complet non orienté sans boucle d'ordre n a une matrice d'adjacence A particulière :

$$a_{ij} = \begin{cases} 1, & \text{pour } i \neq j ; \\ 0, & i = j. \end{cases}$$

Exemple 1.14. La matrice d'adjacence du graphe G_3 de la Figure 1.1 est :

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Théorème 1.1. [179] : Soient un graphe orienté $G = (V, E)$ où $V = \{v_1, v_2, \dots, v_n\}$ et $A = A(G)$ sa matrice d'adjacence. Si $A^k = A \times A \times \dots \times A = a_{ij}^{(k)}$, $k > 1$, alors $a_{ij}^{(k)}$ est le nombre de chemins de longueurs k du sommet v_i au sommet v_j dans G .

Démonstration. Par récurrence. Pour $k = 1$, $a_{ij}^{(1)} = a_{ij}$ est bien le nombre de chemins de longueur 1 de v_i à v_j . Supposons l'énoncé vrai pour $k - 1$. On a $A^k = A^{k-1} \times A$, on a alors

$$A_{ij}^k = \sum_{l=1}^n a_{il}^{(k-1)} \cdot a_{lj}$$

Chapitre 1. Les concepts fondamentaux

et $a_{il}^{(k-1)}$ ni que le nombre de chemins de longueur $k - 1$ de v_i à v_l et

$$a_{lj} = \begin{cases} 1, & \text{si } (v_l, v_j) \in E; \\ 0, & \text{sinon.} \end{cases}$$

On a donc $a_{ij}^{(k)} = \sum_{l=1}^n$ (nombre de chemins de longueur k de v_i à v_j et se terminant par (v_l, v_j)) = (nombre de chemins de longueur k de v_i à v_j). \square

Remarque 1.4. Lorsque $G = (V, E)$ est un graphe simple non orienté où $V = \{v_1, v_2, \dots, v_n\}$, la matrice d'adjacence de G est la matrice d'adjacence du graphe orienté G' associé à G . Cette matrice A est, bien sûr, symétrique. De plus les entrées de A^k comptent le nombre de chaînes de longueur k dans G .

2.2 Matrice de distances

Définition 1.26. Considérons un graphe $G = (V, E)$ d'ordre n . Sa matrice de distance notée $D = (d_{ij})$ est une matrice de dimension $n \times n$, tel que

$$d_{ij} = d(v_i, v_j) \tag{1.3}$$

Où $d(v_i, v_j)$ est la distance entre les deux sommets v_i et v_j .

Exemple 1.15. La matrice de distance du graphe G_3 de la Figure 1.1 est :

$$D = \begin{pmatrix} 0 & 1 & 2 & 3 & 3 & 3 & 2 & 2 & 2 & 2 & 1 \\ 1 & 0 & 1 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 0 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ 3 & 2 & 1 & 0 & 2 & 2 & 3 & 3 & 3 & 3 & 3 \\ 3 & 2 & 1 & 2 & 0 & 2 & 3 & 3 & 3 & 3 & 3 \\ 3 & 2 & 1 & 2 & 2 & 0 & 3 & 3 & 3 & 3 & 3 \\ 2 & 1 & 2 & 3 & 3 & 3 & 0 & 1 & 2 & 2 & 2 \\ 2 & 1 & 2 & 3 & 3 & 3 & 1 & 0 & 2 & 2 & 2 \\ 2 & 1 & 2 & 3 & 3 & 3 & 2 & 2 & 0 & 1 & 2 \\ 2 & 1 & 2 & 3 & 3 & 3 & 2 & 2 & 1 & 0 & 2 \\ 1 & 1 & 2 & 3 & 3 & 3 & 2 & 2 & 2 & 2 & 0 \end{pmatrix}$$

Remarque 1.5. La matrice de distance D d'un graphe G dépend toujours de celle de l'adjacence A du même graphe.

1. La matrice de distance D d'un graphe G peut être construite en calculant successivement les puissances booléennes A^k ; avec $k \in [1, D(G)]$, où $A = I + A$, I soit la matrice unitaire correspond au graphe G et $D(G)$ son diamètre.
2. Dans le cas des graphes orientés, cet algorithme s'appelle l'algorithme de Moore.

2.3 Matrice d'incidence

Définition 1.27. Considérons un graphe orienté sans boucle $G = (V, E)$ d'ordre n et de taille m . Sa matrice d'incidence notée $I = (i_{ij})$ est une matrice de dimension $n \times m$, tel que :

$$i_{ij} = \begin{cases} 1, & \text{si } v_i \text{ est l'extrémité initiale de } e_j ; \\ -1, & \text{si } v_i \text{ est l'extrémité terminale de } e_j ; \\ 0, & \text{si } v_i \text{ n'est pas une extrémité de } e_j. \end{cases}$$

Où v_i sont les sommets de G et e_j sont ses arêtes.

Remarque 1.6. Pour un graphe non orienté sans boucle, la matrice d'incidence est définie par :

$$i_{ij} = \begin{cases} 1, & \text{si } v_i \text{ est une extrémité de } e_j ; \\ 0, & \text{sinon.} \end{cases}$$

Exemple 1.16. La matrice d'incidence du graphe G_2 de la Figure 1.2 est :

$$I = \begin{pmatrix} 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 & 1 & 0 \\ -1 & -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

2.4 Listes d'adjacences

Le sommet	ses successeurs	ses prédécesseurs
u_1	u_5	u_4
u_2	u_4	u_5
u_3	u_5	u_4
u_4	u_1, u_3	u_2
u_5	u_2	u_1, u_3

Tableau 1.1 – Les listes d'adjacences du graphe G_2 de la Figure 1.2

Chapitre 1. Les concepts fondamentaux

Un graphe orienté peut être représenté à l'aide d'un dictionnaire ; il s'agit d'une table à simple entrée où chaque ligne correspond à un sommet et comporte la liste des successeurs ou des prédécesseurs de ce sommet. Dans la mesure où, pour une table donnée, le nombre de successeurs ou de prédécesseurs n'est pas le même pour chaque sommet, il est préférable de représenter le dictionnaire sous forme de deux tableaux : le premier comprenant autant d'éléments que de sommets, ces éléments pointant, dans un second tableau, les débuts de listes de successeurs (ou de prédécesseurs).

Exemple 1.17. Le tableau 1.1 affiche les listes d'adjacences du graphe G_2 de la Figure 1.2.

2.5 Matrice Laplacienne

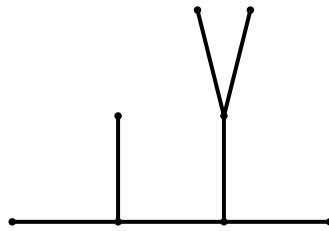


Figure 1.3 – Exemple d'un arbre

Définition 1.28. La matrice Laplacienne notée $L(G)$ correspondant au graphe G est la matrice définie comme suit :

$$L(G) = D(G) - A(G) \quad (1.4)$$

Où $D(G)$ est la matrice diagonale comportant les degrés des sommets du graphe, et $A(G)$ sa matrice d'adjacence.

Exemple 1.18. La matrice Laplacienne du graphe de la Figure 1.3 est :

$$L(G) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 3 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Remarque 1.7. Par la suite ne nous concéderons que les graphes simples, non orientés, plainaires, connexes et finis.

3 La théorie des graphes moléculaires

La chimie est qualifiée de «science centrale» en raison des puissants liens qu'elle possède avec les mathématiques, la biologie et la physique, ainsi qu'avec la médecine, la pharmacie, l'informatique et la science des matériaux, sans oublier des domaines appliqués tels que le génie des procédés. Nous nous intéresserons à sa relation avec l'informatique, et pour bien comprendre l'interprétation chimique de nos travaux, nous seront obligés d'introduire quelques notions de base de chimie, qui vont être utiles dans la suite de notre travail.

3.1 Les concepts généraux

La chimie est une science de la nature. Elle est divisée en plusieurs spécialités expérimentales et théoriques à l'instar de la physique et de la biologie avec lesquelles elle partage parfois des espaces d'investigations communs ou proches. Selon l'American Chemical Society (ACS), une grande association de chimistes au monde, la chimie étudie :

- Les éléments chimiques à l'état libre, atomes ou ions atomiques, et les innombrables et diverses associations par liaisons chimiques qui engendrent notamment des composés moléculaires stables ou des intermédiaires plus ou moins instables. Ces entités de matière peuvent être caractérisées par une identité reliée à des caractéristiques quantiques et des propriétés précises ;
- Les processus qui changent ou modifient l'identité de ces particules ou molécules de matière, dénommés réaction, transformation, interaction... ;
- Les mécanismes intervenant dans les processus chimiques ou les équilibres physiques entre deux formes. Leurs définitions précises permettent de comprendre ou d'interpréter avec des hypothèses l'évolution matérielle avec en vue une exploitation des résultats de façon directe ou induite ;
- Les phénomènes fondamentaux observables en rapport avec les forces de la nature qui jouent un rôle chimique, favorisant les réactions ou synthèse, addition, combinaison ou décomposition, séparation de phases ou extraction. L'analyse permet de découvrir les compositions, le marquage sélectif ouvre la voie à un schéma réactionnel cohérent dans des mélanges complexes.

Chapitre 1. Les concepts fondamentaux

Définition 1.29. La matière est ce qui compose tout corps ayant une réalité tangible (que l'on peut connaître en touchant).

Définition 1.30. Un élément chimique désigne l'ensemble des atomes caractérisés par un nombre défini de protons dans leur noyau atomique. L'hydrogène, le carbone, l'azote, l'oxygène, sont des éléments chimiques, de même que le fer, le cuivre, l'argent, l'or, etc. Chacun est conventionnellement désigné par un symbole chimique, tel que H , C , N , O , ou encore Fe , Cu , Ag , Au , etc. Au total, 118 éléments chimiques ont été observés à ce jour. Parmi ceux-ci, 94 éléments se rencontrent dans le milieu naturel. Les éléments chimiques peuvent se combiner entre eux au cours de réactions chimiques pour former d'innombrables composés chimiques

Définition 1.31. Un corps simple est une forme chimique ne comportant qu'un élément chimique, par opposition aux composés chimiques, infiniment plus nombreux et dans lesquels plusieurs éléments différents sont associés de diverses manières.

Définition 1.32. Un atome est la plus petite partie d'un corps simple pouvant se combiner chimiquement avec une autre. Il est généralement constitué d'un noyau composé de protons et de neutrons autour desquels orbitent des électrons. Sa taille caractéristique est de l'ordre du dixième de nanomètre (nm).

Définition 1.33. La liaison chimique est le phénomène physique qui lie les atomes entre eux en échangeant ou partageant un ou plusieurs électrons ou par des forces électrostatiques.

Définition 1.34. Une molécule est une structure de base de la matière. C'est l'ensemble d'atomes unis les uns aux autres par des liaisons chimiques. La composition chimique d'une molécule est donnée par sa formule chimique. Exemples :

- La molécule de méthane CH_4 est constituée d'un atome de carbone(C) et de quatre atomes d'hydrogène(H) ;
- La molécule de dioxygène O_2 est constituée de deux atomes d'oxygène (O).

Définition 1.35. La formule chimique (structurée) est la représentation de molécules qui s'utilise en chimie pour décrire les molécules et leurs structures, cette représentation graphique permet de décrire les liaisons moléculaires, le nombre et le type d'atomes qui composent une molécule, sa forme dans l'espace ou simplement de décrire sommairement la molécule de manière simple et rapide. La plupart de ces représentations sont surtout utilisées en chimie organique ou en biochimie.

Exemple 1.19. Les formules chimiques de 2,2,4,6--tetraméthylheptane (I) et diphenylène (II) respectivement figurent dans la Figure 1.4.

3.2 Les graphes moléculaires

Dans le jargon mathématique on appelle le thème traitant les graphes, la théorie des graphes. Comme nous l'avons vu au début de ce chapitre, un graphe est un objet mathématique identifié essentiellement par ses deux principaux éléments : arête et sommet. Dans un graphe toute arête correspond à deux sommets, appelés sommets adjacents qui sont habituellement dessinés par des petits cercles ou des grands points, et les arêtes sont dessinées par des traits.

Aujourd'hui, la théorie des graphes a connu plusieurs applications dans divers domaines, tels que l'électro-technique, la sociologie, la physique nucléaire, l'informatique, l'ethnologie, l'ingénierie, la géographie, la linguistique, la biologie, le transport et en particulier la chimie organique. Nombreux livres [17, 24, 38, 40, 118, 137, 164, 169] et articles [27, 46, 173, 175, 202] ont été écrits, montrent les applications des graphes en chimie. Ces applications chimiques sont basées sur le fait qu'une similarité existe entre un composé chimique et un graphe. La Figure 1.4, montre les composés chimiques de 2, 2, 4, 6--tetramethylheptane (I) et diphenylene (II) respectivement, qui ressemblent au niveau de formes aux graphes G_1 et G_2 de la Figure 1.1. Évidemment G_1 et G_2 peuvent être vus comme les représentations graphiques des formules chimiques des molécules 2, 2, 4, 6--tetramethylheptane et diphenylene respectivement. Donc on dit que G_1 est le graphe chimique de 2, 2, 4, 6--tetramethylheptane et le graphe G_2 et celui de diphenylene.

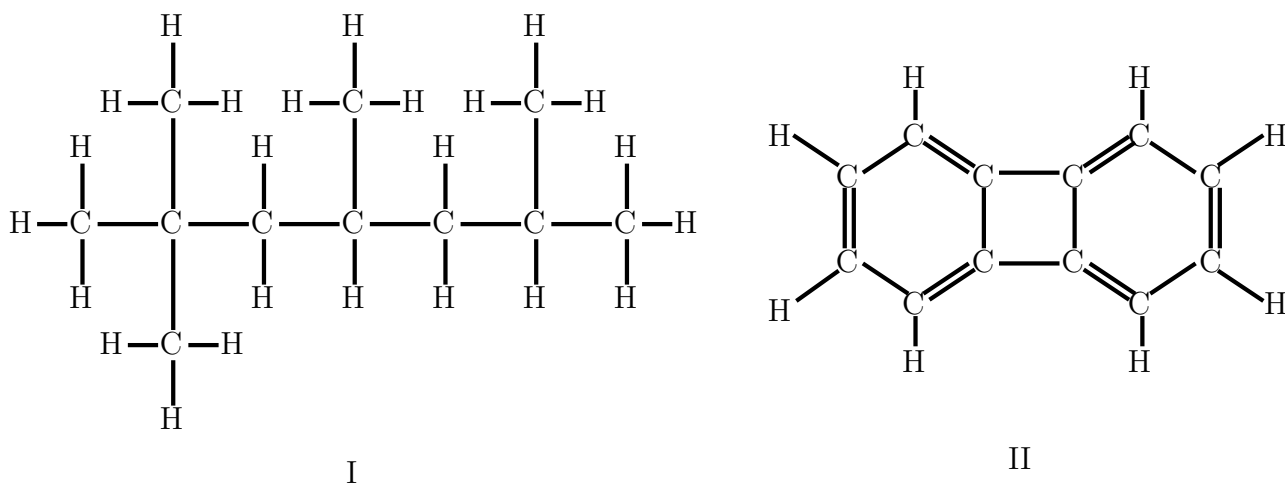


Figure 1.4 – les composés chimiques de 2, 2, 4, 6--tetramethylheptane (I) et diphenylene (II)

Remarque 1.8. Notant que :

- Les atomes d'hydrogène et les liaisons doubles sont supprimés dans G_1 et G_2 , ce n'est pas nécessaire, mais il a été prouvé que c'est demandé dans la pratique, voyez l'article [117], pour plus de détails.

- Que les graphes chimiques sont obligatoirement connectés [41, 119, 170].
- Ils y a des graphes qui ne sont pas des graphes chimiques, par exemple le graphe G_3 de la Figure 1.1.

3.2.1 Définition

On appelle un graphe chimique ou un graphe moléculaire tout modèle d'un système chimique, qu'on utilise pour caractériser les interactions entre ses composantes chimiques : les atomes, les liaisons, les groupes d'atomes ou bien de molécules. La formule structurale d'un composé

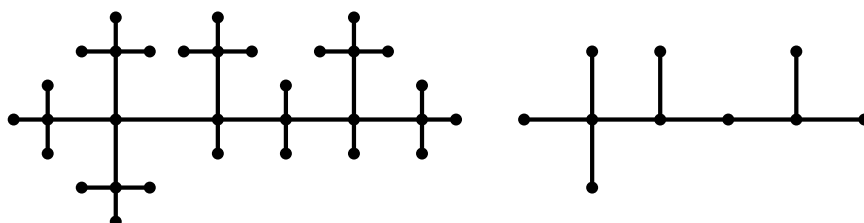


Figure 1.5 – Les représentations Plerograph (Pl) et kenograph (Ke) de 2, 2, 3, 5 – *tetramethylhexane*

chimique peut être représentée par un graphe moléculaire, dont ses sommets représentent les atomes et ses arêtes correspondent aux liaisons chimiques.

3.2.2 La représentation d'une molécule

La notion d'un graphe chimique a été utilisée depuis le dix-neuvième siècle par plusieurs chimistes. Mr Arthur Cayley s'est considéré parmi les inventeurs de ce concept, ce qui est clair dans ses articles [9][167]. Dans son article «chemical graph theory» [9] Cayley a proposé deux représentations graphiques d'une molécule, la représentation «**plerogram**» et celle «**kenogram**», définissent ainsi :

Définition 1.36. La représentation **Plerogram** est le graphe moléculaire dans lequel tous les atomes de la molécule y compris ceux d'hydrogènes sont représentés par des sommets, et les liaisons entre eux sont représentées par des arêtes.

Définition 1.37. La représentation **Kenogram** est le graphe moléculaire où on ne représente par des sommets que les atomes des carbones d'une molécule. On l'appelle aussi le **graphe à atome d'hydrogène-supprimé**, la Figure 1.5 montre les deux représentations de 2, 2, 3, 5 – *tetramethylhexane* [163].

Chapitre 2

La prédiction de propriétés et activités moléculaires

L'utilisation des outils informatiques chez les chimistes est devenue obligatoire afin de bien manipuler les informations moléculaires qui ont été, au cours des dernières années, stockées numériquement sur les ordinateurs dans des bases de données en très grandes quantités. De plus, la multiplication des données exploitables par les chimistes a donné lieu à une obligation de la numérisation, afin d'être capable de stocker, visualiser et traiter ces mêmes données aisément.

La discipline décrivant l'utilisation des outils informatiques pour traiter et résoudre des problèmes à la fois dans le domaine chimique et biologique est désignée par «**La Chémoinformatique**». Ses utilisations sont très variées et vont de la création et l'utilisation de base de données de petites molécules à la manipulation de fichiers en passant par les études statistiques. Cependant, son application la plus communément admise est dans le domaine de la recherche de nouveaux médicaments «**drug discovery**», domaine dans lequel elle joue un rôle central dans l'analyse et l'interprétation des données de structures et de propriétés collectées au cours des criblages à haut débit (technique se fait par les biologiste et visant à identifier des molécules nouvelles et potentiellement actives dans des bases de données de composés).

L'émergence de cette discipline peut être mise en parallèle avec la multiplication des données chimiques stockées numériquement. En effet, les quantités de données générées par les nouvelles approches de drug design n'ont eu de cesse d'augmenter et il s'est avéré nécessaire, pour traiter les résultats de criblage à haut débit ou encore de la chimie combinatoire, de développer et d'utiliser des techniques informatiques [69].

Les avancées technologiques de la dernière décennie ont rendu possibles de nombreuses découvertes et applications inaccessibles auparavant. Par exemple, le nombre de composés disponibles dans les études de criblage a augmenté de manière exponentielle. En parallèle, les développe-

ments techniques dans le domaine de l'informatique et des technologies de communication ont permis la création de bases de données de composés comportant des millions d'entrées. L'exemple le plus pertinent pour illustrer ces avancées est la base de données «**PubChem**» développée par le *NIH* [63]. Avec un contenu de plus de 31 millions de composés reliés à leurs activités biologiques, ce genre de bases de données nécessite le développement et l'utilisation d'outils mathématiques et statistiques afin de pouvoir accéder à de nouvelles découvertes en termes de développement de nouveaux médicaments et à la compréhension «**des relations entre structure et activité (QSAR)**» ou bien «**des relations entre structure et propriétés (QSPR)**».

Ce présent chapitre vise explorer les principales disciplines dans lesquelles on manipule les indices topologiques telle que les méthodes *QSAR/QSPR* et le criblage virtuel. On commence tout d'abord par la définition d'un indice topologique, ses différentes catégories et quelques exemples de chacune de ses catégories. Nous introduirons ensuite globalement et d'une façon rapide la définition de *QSAR/QSPR*, ses étapes et ses principales méthodes de modélisation exploitées en *QSAR/QSPR*. On termine ce chapitre par une introduction rapide au criblage à haut débit, mais cette fois-ci les tests seront faits par les informaticiens et n'est pas dans les laboratoires; c'est-à-dire : le criblage virtuel.

1 Les descripteurs moléculaires

De nombreuses recherches ont été menées, au cours des dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules, et ces structures elles-mêmes, en un ensemble de nombres réels appelés descripteurs; une fois que ces nombres sont disponibles, il est possible d'établir une relation entre ceux-ci et une propriété ou activité moléculaire, à l'aide d'outils de modélisation classiques.

1.1 Définition

Les descripteurs numériques réalisent un codage de l'information chimique en un vecteur de réels. Tout simplement, un descripteur moléculaire est une représentation mathématique d'une molécule, qui contient à la fois des informations sur la structure, et donc, implicitement ou explicitement, sur ses propriétés physico-chimiques. Ces informations peuvent être encodées par des valeurs scalaires, des vecteurs ou des chaînes de bits [16, 70, 163].

1.2 Les types de descripteurs moléculaires

On dénombre aujourd’hui plus de 10000 descripteurs moléculaires, qui quantifient des caractéristiques physico-chimiques ou structurelles de molécules. Ils peuvent être obtenus de manière **empirique** ou **non-empirique**, mais les descripteurs calculés, et non mesurés, sont à privilégier : ils permettent en effet d’effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs de la modélisation. Il existe cependant quelques descripteurs mesurés : il s’agit généralement de données expérimentales plus faciles à mesurer que la propriété ou l’activité à prédire (coefficient de partage eau-octanol [36], polarisabilité, ou potentiel d’ionisation). Les descripteurs moléculaires sont fréquemment classés par rapport à la dimensionnalité de la représentation moléculaire sur laquelle ils sont calculés : On parlera alors de **descripteurs 1D**, **2D**, ou **3D** [28].

1.2.1 Les descripteurs 1D

Les descripteurs 1D sont appelés «**descripteurs constitutionnels**» et sont faciles et rapides à calculer. Ils sont accessibles à partir de la formule brute de la molécule (par exemple C_6H_6O pour le phénol), et décrivent des propriétés globales du composé. Il s’agit par exemple de sa composition, c’est-à-dire les atomes qui le constituent, ou de sa masse molaire. On peut remarquer que ces descripteurs ne permettent pas de distinguer les isomères [16, 70].

1.2.2 Les descripteurs 2D

Les descripteurs moléculaires utilisent la représentation des molécules comme des graphes sont dits «**descripteurs 2D**» et contiennent des informations à propos de la connectivité ou à propos de certains fragments moléculaires, mais aussi des estimations des propriétés physico-chimiques. C’est à partir de ce niveau que l’on peut espérer la capture d’informations chimiques pertinentes pour la prédiction de la majorité des propriétés moléculaires. On trouvera dans cette catégorie les descripteurs suivants :

- **Les indices topologiques**, qui considèrent la structure du composé comme un graphe, les atomes étant les sommets et les liaisons sont les arêtes. De nombreux indices quantifiant la connectivité moléculaire ont été développés en se basant sur cette approche, comme par exemple l’**indice de Wiener** [98], qui compte le nombre total de liaisons dans les chemins les plus courts entre toutes les paires d’atomes (en excluant les hydrogènes), et qui sera également l’axe central de cette thèse. D’autres indices basés sur les chemins ont été développés [25, 84, 153].

Le but de cette thèse, donc tour au tour de cette catégorie, l'étude de ses principales indices, de différentes façons de leurs calculs et de la présentation de leurs fondamentales propriétés. Nous étudierons, donc, dans les prochains chapitres les indices de Wiener, de Hyper-Wiener, de degré de distance, Terminal de Wiener, en démontrons les principales corrélations qui peuvent exister entre eux et d'autres indices de la même catégorie. Comme nous allons étudier les deux principales représentations moléculaires sur lesquelles on se base pour calculer les indices de cette catégorie.

Les informations sur les électrons de valence peuvent être incluses dans les indices topologiques [131, 143]. Enfin, des descripteurs combinant les informations de connectivité avec d'autres propriétés sont aussi à disposition, comme par exemple les descripteurs BCUT, qui se présentent sous la forme de matrices de connectivités des atomes, avec sur la diagonale la charge atomique, la polarisabilité ou les valeurs du potentiel de liaisons hydrogènes, et des termes additionnels hors diagonale [59, 72].

- **Les indices constitutionnels**, qui se basent sur des motifs sous-structuraux. Par exemple, les empreintes *BCI* [135] sont des ensembles de bits indiquant la présence ou l'absence de certains fragments dans une molécule. Les fragments prennent en compte les atomes et leurs plus proches voisins, les paires d'atomes et les séquences ou encore les fragments basés sur des cycles. L'approche des clés MDL est une approche similaire comprenant la recherche des 166 fragments *MDL* [16, 70, 134].

Ces descripteurs *2D* reflètent bien les propriétés physiques dans la plupart des cas, mais sont insuffisants pour expliquer de façon satisfaisante certaines propriétés ou activités, telles que les activités biologiques. Des descripteurs, accessibles à partir de la structure *3D* des molécules, ont pu être calculés grâce au développement des techniques instrumentales et de nouvelles méthodes théoriques.

1.2.3 Les descripteurs *3D*

«**Les descripteurs *3D***» d'une molécule sont évalués à partir des positions relatives de ses atomes dans l'espace, et décrivent des caractéristiques plus complexes ; leurs calculs nécessitent donc de connaître, le plus souvent par «**modélisation moléculaire empirique**» ou «**ab initio**», la géométrie *3D* de la molécule. Ces descripteurs s'avèrent donc relativement coûteux en temps de calcul, mais apportent davantage d'informations, et sont nécessaires à la modélisation de propriétés ou d'activités qui dépendent de la structure *3D*. On distingue plusieurs familles importantes de descripteurs *3D* :

- **Les descripteurs géométriques**, parmi ceux qui sont les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d’inertie.
- **Les descripteurs électroniques**, ils permettent de quantifier différents types d’interactions inter- et intra-moléculaires, de grande influence sur l’activité biologique de molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l’énergie stérique est minimale, et fait souvent appel à la chimie quantique. Par exemple, les énergies de la plus haute orbitale moléculaire occupée et de la plus basse vacante sont des descripteurs fréquemment sélectionnés. Le moment dipolaire, le potentiel d’ionisation, et différentes énergies relatives à la molécule sont d’autres paramètres importants.
- **Les descripteurs spectroscopiques** : les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d’onde vibrationnelles. En effet, les vibrations d’une molécule dépendent de la masse des atomes et des forces d’interaction entre ceux-ci ; ces vibrations fournissent donc des informations sur la structure de la molécule et sur sa conformation. Les spectres infrarouges peuvent être obtenus soit de manière expérimentale, soit par calcul théorique, après recherche de la géométrie optimale de la molécule. Ces spectres sont alors codés en vecteurs de descripteurs de taille fixe. Le descripteur *EVA* [194] est ainsi obtenu à partir des fréquences de vibration de chaque molécule. Les descripteurs de type MoRSE [132] (Molecule Representation of Structures based on Electron diffraction) sont calculés à partir d’une simulation du spectre infrarouge ; ils font appel au calcul des intensités théoriques de diffraction d’électrons.

Ces descripteurs requièrent une conformation 3D de la molécule expérimentale ou prédite. De plus, l’information sur la structure de la cible (protéine) est parfois requise. On pourra ainsi distinguer les descripteurs 3D qui nécessitent un alignement de la molécule guidé par l’étude des complexes ligand-cible (ou, au moins, par des contraintes visant d’optimiser le recouvrement spatial des champs électriques et stériques des ligands, faute d’information sur le vrai mode de fixation dans la cible) avant d’être calculés, comme par exemple les descripteurs CoMFA [16, 70, 177].

2 Les études QSAR/QSPR

Les premiers essais de modélisation d’activités de molécules datent de la fin du 19^{ème} siècle, lorsque Crum-Brown et Frazer [12] postulèrent que l’activité biologique d’une molécule est une fonction de sa constitution chimique. Mais ce n’est qu’en 1964 que furent développés les modèles de «contribution de groupes», qui constituent les réels débuts de la modélisation QSAR. Depuis, l’essor de nouvelles techniques de modélisation par apprentissage, linéaires d’abord, puis non

linéaires, ont permis la mise en place de nombreuses méthodes ; elles reposent pour la plupart sur «la recherche d'une relation entre un ensemble de nombres réels, descripteurs de la molécule, et la propriété ou l'activité que l'on souhaite prédire».

2.1 Définition de QSAR/QSPR

Lors d'une étude de «**QSAR (Qualitative Structure-Activity Relationships)**» ou de «**QSPR (Qualitative Structure-property Relationships)**», on étudie les relations entre la structure et l'activité (propriété) d'un composé ou molécule, par exemple les effets d'une variation chimique locale sur une molécule à l'activité connue. En effet, certains changements chimiques sur certaines parties d'une molécule peuvent entraîner des variations de son activité biologique en agissant sur l'interaction avec la cible [28].

Ainsi, la méthodologie **QSAR/QSPR** permet de trouver un modèle mathématique qui mette en corrélation l'activité (propriété) et la structure au sein d'une famille de composés. De nombreuses méthodes conceptuellement différentes peuvent être utilisées pour mettre en place les modèles mathématiques permettant de détecter des relations de type **QSAR**. Ainsi, les études **QSAR/QSPR** sont basées sur des méthodes informatiques, celles de modélisations, déjà exploitées dans différents domaines.

Les grandes phases de la mise en place d'un modèle **QSAR/QSPR** peuvent être décrites comme suit : Extraire les descripteurs à partir de la structure moléculaire, choisir les descripteurs adaptés à l'étude par rapport à l'activité (propriété) analysée, utiliser les descripteurs comme variables explicatives pour définir une relation qui les corrèle à l'activité en question, et enfin chaque modèle doit être validé sur des jeux de données de test [70].

2.2 Préparation des données d'entrée

Une bonne préparation des données d'entrée est nécessaire pour une étude **QSAR/QSPR** efficace. Cette préparation implique plusieurs étapes de travail sur les composés :

- S'assurer que les conditions expérimentales dans lesquelles ont été obtenues les mesures d'activité des molécules sont similaires,
- éliminer les doublons,
- Appliquer les mêmes règles de standardisation pour les structures des composés,
- éliminer les mélanges (sauf si, bien sûr, il faut prédire les propriétés de mélanges).

2.3 Génération des descripteurs moléculaires

Une fois les données d'entrée sont préparées, la première étape de l'étude consiste à obtenir un ensemble de descripteurs pour chacune des molécules.

En effet, les composés, encodés comme des ensembles de liaisons covalentes et d'atomes, ne peuvent être utilisés directement sur un ordinateur. Les structures chimiques ne contiennent pas habituellement d'information explicite les reliant à l'activité. Cette information doit être extraite grâce au descripteurs variés qu'il est possible de mettre en place. De cette manière, des propriétés implicites contenues dans la structure de la molécule vont être mises en avant, sachant que seules certaines d'entre elles peuvent éventuellement corrélérer avec l'activité (la propriété). Comme indiqué au paragraphe des descripteurs moléculaire (paragraphe I de ce chapitre), ces descripteurs sont basés non seulement sur la structure des composés mais aussi sur un ensemble de propriétés physico-chimiques.

Pour des raisons techniques, les descripteurs sont représentés sous la forme de vecteurs de nombres pour chaque molécule. Ces vecteurs doivent tous être de la même longueur. En effet, la plupart des méthodes utilisées pour la prédiction requièrent comme données d'entrée des objets comparables de taille constante, ce qui n'est pas naturellement le cas des structures des composés, qui sont diverses en taille et en nature.

2.4 Sélection des descripteurs les plus adaptés

Avant toute modélisation, il est nécessaire de calculer ou de mesurer un grand nombre de descripteurs différents, car les mécanismes qui déterminent l'activité d'une molécule ou une de ses propriétés sont fréquemment mal connus. Cependant, les descripteurs envisagés n'ont pas tous une influence significative sur la grandeur modélisée, et les variables ne sont pas toujours mutuellement indépendantes. De plus, le nombre de descripteurs, c'est-à-dire la dimension du vecteur d'entrée, détermine la dimension du vecteur des paramètres à ajuster. Si cette dimension est trop importante par rapport au nombre d'exemples de la base d'apprentissage, le modèle risque d'être surajusté à ces exemples, et incapable de prédire la grandeur (activité ou propriété) modélisée sur de nouvelles observations. L'ensemble de descripteurs doit donc être le plus petit possible mais le plus riche en informations possible.

Pour cette raison, des méthodes spécifiques doivent être utilisées afin de réduire le nombre de descripteurs aux descripteurs les plus informatifs. Plusieurs approches sont possibles pour résoudre ce problème :

- Réduire la dimension de l'espace des entrées,

- Remplacer les variables corrélées par de nouvelles variables synthétiques, obtenues à partir de leurs combinaisons,
- Sélectionner les variables les plus pertinentes.

Nous allons maintenant décrire brièvement les méthodes les plus fréquemment utilisées.

2.4.1 L'analyse en composantes principales

L'analyse en composantes principales (ou *ACP*) [128], est une technique d'analyse de données utilisée pour réduire la dimension de l'espace de représentation des données. Contrairement à d'autres méthodes de sélection, celle-ci porte uniquement sur les variables, indépendamment des grandeurs (propriété ou activité) que l'on cherche à modéliser. Les variables initiales sont remplacées par de nouvelles variables, appelées composantes principales, deux à deux non corrélées, et telles que les projections des données sur ces composantes soient de variance maximale. Elles peuvent être classées par ordre d'importance. Puisque les composantes principales sont des combinaisons linéaires des variables initiales, l'interprétation du rôle de chacune de ces composantes reste possible. Il suffit en effet de déterminer quels descripteurs d'origine leur sont le plus fortement corrélés. Les variables obtenues peuvent ensuite être utilisées en tant que nouvelles variables du modèle. Par exemple, la régression sur composantes principales [82] (ou *PCR*) est une méthode de modélisation dont la première étape est une analyse en composantes principales, suivie d'une régression linéaire multiple (voir le prochain paragraphe).

2.4.2 La méthode de régression des moindres carrés partiels

La régression des moindres carrés partiels [21, 101] (*MCP*, ou *PLS*) est également une méthode statistique utilisée pour construire des modèles prédictifs lorsque le nombre de variables est élevé et que celles-ci sont fortement corrélées. Cette méthode utilise à la fois des principes de l'*ACP* et de la régression multilinéaire. Elle consiste à remplacer l'espace initial des variables par un espace de plus faible dimension, sous-tendu par un petit nombre de variables appelées «**variable latentes**», construites de façon itérative. Les variables retenues sont orthogonales (non corrélées), et sont des combinaisons linéaires des variables initiales. Les variables latentes sont obtenues à partir des variables initiales, mais en tenant compte de leur corrélation avec la variable (activité ou propriété) modélisée, contrairement aux variables résultant de l'analyse en composantes principales. Elles doivent ainsi expliquer le mieux possible la covariance entre les entrées et la sortie. Elles sont alors les nouvelles variables explicatives d'un modèle de régression classique, telles que la régression linéaire multiple.

2.4.3 sélection de variables pertinentes

L'analyse en composantes principales a pour but de réduire les corrélations entre les variables, mais cette étape de réduction est indépendante de la grandeur modélisée. Or, les variables calculées n'ont pas nécessairement une influence sur cette grandeur. Il est nécessaire d'éliminer celles dont l'influence est inférieure à celle du bruit, et de sélectionner uniquement les plus pertinentes d'entre elles. Cette méthode permet d'envisager toutes les combinaisons possibles de descripteurs, mais sa mise en oeuvre est très lourde. Des approches alternatives ont donc été développées :

- *La sélection progressive*, qui consiste à incorporer les variables au modèle une à une, en sélectionnant, à chaque étape, la variable dont la corrélation partielle avec la grandeur modélisée est la plus élevée. À l'inverse, lors de l'élimination progressive, on débute la modélisation avec l'ensemble des descripteurs, en les éliminant un par un jusqu'à obtenir le meilleur jeu de composantes.
- *La sélection pas à pas*, qui est une combinaison des deux méthodes évoquées précédemment. Les variables sont incorporées une à une dans le modèle, par sélection progressive. Cependant, à chaque étape, on vérifie que les corrélations partielles des variables précédemment introduites sont encore significatives.
- *D'autres méthodes plus efficaces* à ces deux dont nous venons de les annoncer existent par exemple : la méthode du descripteur sonde, pour plus de détail voir [92, 94].

Dans le suivant paragraphe nous donnerons l'idée de notre contribution dans les techniques de sélection de variables.

2.4.4 Le regroupement de plusieurs indices

En voyant ces techniques de la sélection des variables, on peut les relier à quelques travaux qu'on a pu le faire durant la durée de cette thèse. Ces travaux portent sur le regroupement de plusieurs indices topologiques en une seule définition. Donc, on a arrivé à donner une définition unifiée aux indices topologiques utilisant la distance, les degrés, ... en leurs définitions (voir le dernier chapitre).

L'idée de regroupement de plusieurs définitions des indices en une seule est presque la même chose que l'approche «**Remplacer les variables corrélées par de nouvelles variables synthétiques, obtenues à partir de leurs combinaisons**», et donne le même résultat que la technique **d'analyse en composantes principales** vu auparavant, sauf que pour nous, on les remplace au début des étapes des études *QSAR/QSPR*, par contre les méthodes de sélection de variables vu dans cette sous section les réduisent au cours de traitement et en tant

que possible. Donc, théoriquement on peut les regrouper en une seule définition ce qui facilite leur utilisation aux méthodes de *QSAR/QSPR*, et pratiquement, nous les rendrons parmi nos érgents perspectives. Nous allons arriver à ces regroupements, qui seront expliqués en détail dans le chapitre V de ce mémoire.

2.5 Relation des descripteurs à la grandeur

Une fois les descripteurs moléculaires utiles calculés et sélectionnés, l'avant dernière étape est de créer une fonction reliant leurs valeurs à la grandeur analysée. La valeur qui quantifie la grandeur (activité ou propriété) sera donc exprimée comme étant une fonction des valeurs des descripteurs. Les meilleures fonctions sont en général mises en place en se basant sur l'information contenue dans l'ensemble d'entraînement (les composés pour lesquels la grandeur est connue).

Comme il est expliqué en Annexe A, il existe une gamme de familles de fonctions très vaste, incluant des «**fonctions linéaires**», qui ont été utilisées depuis le début du **QSAR**. Ces fonctions prédisent la grandeur comme étant une fonction linéaire des descripteurs moléculaires. En général, ces fonctions linéaires sont facilement interprétables et suffisamment précises pour de petits ensembles de composés similaires, spécialement lorsque les descripteurs sont sélectionnés avec soin pour une grandeur donnée. Ces méthodes sont par exemple **la régression linéaire multiple** (*MLR : Multiple Linear Regression*) [16, 37, 189, 193], **La méthode des moindres carrés** (*PLS : Partial Least Squares*) [16, 190, 191] ou encore l'**analyse discriminante linéaire** (*LDA : Linear Discriminant Analysis*) [180].

D'autres méthodes, «**non-linéaires**», étendent l'approche à des relations plus complexes. Ces méthodes génèrent des modèles se révèlent être plus précis, spécialement pour des ensembles de données plus larges et plus divers. Cependant, ces modèles nous heurtent parfois à des difficultés de compréhension et sont parfois en proie à l'**overfitting** (ils se borneront dans ce cas à décrire du bruit au lieu de la relation sous-jacente entre descripteurs et activité). On utilisera comme méthodes non-linéaires **la classification de Bayes** [198], **la méthode des k plus proches voisins** ($k - NN$) [192], **des réseaux de neurones** [16, 22, 73], mais aussi des **arbres de décision** [136, 183], ou des **méthodes des machines à vecteurs de support** (*SVM Vector Machines Support*) [16, 35].

2.6 la validation de moldèle

La modélisation vise à fournir un modèle qui soit non seulement ajusté aux données d'apprentissage, mais aussi capable de prédire la valeur de la sortie sur de nouveaux exemples, c'est-à-dire

de généraliser. La dernière phase des études QSAP/QSPR donc est de l'évaluer pour qu'on serai sur de son utilisation dans la pratique.

Plusieurs techniques on été conçus pour la sélection de modèle, pour en savoir plus, voir [16].

2.7 Autres méthodes de QSAR/QSPR

La modélisation d'une propriété ou d'une activité moléculaire nécessite de disposer d'informations caractérisant les molécules, informations à partir desquelles la grandeur en question est prédite. Il peut s'agir de descripteurs, mais il existe des méthodes alternatives de caractérisation des molécules.

Nous avons vu les utilisations des descripteurs $2D$ et les méthodes qui les utilisent dans ce qui est passé, maintenant nous allons donner deux exemples des utilisations des autres descripteurs, un pour $1D$ et l'autre pour $2D$. Nous présenterons dans un premier temps la méthode de contribution de groupes (exemple pour les descripteurs $1D$), qui, bien que datant des débuts de la modélisation **QSAR**, est toujours utilisée pour des applications particulières. Puis la méthode d'Analyse comparative de champs moléculaires (exemple pour les descripteurs $3D$), qui a montré son efficacité dans la discipline de modélisation.

2.7.1 Méthode de contribution de groupes

«**Les méthodes de contribution de groupes**» consistent à évaluer une propriété en décomposant la molécule en un ensemble de groupes fonctionnels, et en sommant les contributions relatives à des fragments de molécules [23, 76]. Ces contributions sont déterminées à partir d'une base d'exemples de molécules, dont les valeurs de la propriété sont connues. Plusieurs types de groupes fonctionnels peuvent être définis. Ils sont généralement organisés en un système hiérarchique :

- **Les groupes d'ordre 0** sont des atomes, et le calcul d'une propriété est effectué en sommant les contributions de chacun des atomes de la molécule considérée.
- **La décomposition en groupes d'ordre 1** consiste à découper la molécule en groupes d'atomes (tels que $-CH_2-$, $-CH_3$ ou $-OH$). Leurs contributions à une propriété donnée sont sommées sans que l'environnement de chacun des groupes dans la molécule ne soit pris en considération. Ainsi, le groupe $-CH_2-$ a une contribution fixe, qu'il soit relié à un carbone ou à un groupe oxygéné. Ces groupes sont assez souvent employés, car ils permettent d'estimer rapidement la valeur d'une propriété, avec une précision parfois suffisante (par exemple pour l'enthalpie de formation). Cependant, les résultats obtenus, pour la température d'ébullition

Chapitre 2. La prédiction de propriétés et activités moléculaires

par exemple, ne sont pas toujours satisfaisants. De plus, certains isomères peuvent conduire à la même décomposition : il est alors impossible de les distinguer par cette méthode.

- **Les groupes d'ordre 2** [16, 140] sont constitués des atomes centraux de la molécule (autres que H), accompagnés de leurs plus proches voisins, c'est-à-dire de tous les atomes auxquels ils sont reliés. Contrairement aux groupes d'ordre 1, ceux d'ordre 2 tiennent compte de l'environnement des atomes.

Exemple 2.1. Le Tableau 2.1 présente une comparaison des décompositions des molécules de *butan-2-ol* et *2-méthylpropan-1-ol*, représentées sur la Figure 2.1. On observe que

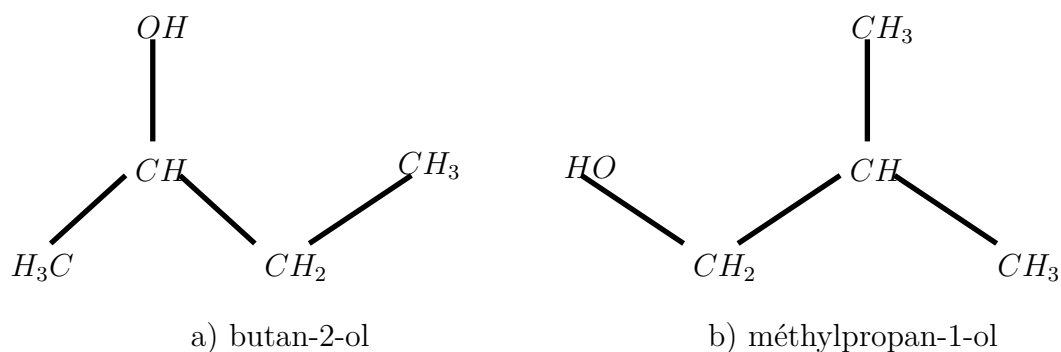


Figure 2.1 – Exemple de deux molécules, isomères de constitution

seule la décomposition en groupes d'ordre 2 permet de distinguer ces deux molécules.

Méthode	Groupe	Nombre de groupe a)	Nombre de groupe b)
Groupe 0	C	4	4
	H	10	10
	O	1	1
Groupe 1	CH	1	1
	$-CH_2-$	1	1
	$-CH_3$	2	2
	$-OH$	1	1
Groupe 2	$C - (C)(H)_3$	2	2
	$C - (C)_2(H)_2$	1	0
	$C - (C)(O)(H)_2$	0	1
	$C - (C)_3(H)$	0	1
	$C - (C)_2(O)(H)$	1	0

Tableau 2.1 – Décomposition en groupes de deux isomères de constitution

De nombreuses méthodes s'appuient donc sur des groupes des trois ordres pour améliorer la précision des prédictions et différencier les isomères. Elles sont principalement utilisées pour prédire des propriétés thermodynamiques, par exemples des propriétés critiques (température ou pression critique) et de nombreuses grandeurs énergétiques. Elles présentent également l'avantage de permettre l'estimation de propriétés de mélanges, par addition des contributions des composants du mélange.

2.7.2 Analyse comparative de champs moléculaires (CoMFA)

«**La méthode *CoMFA* (Comparative Molecular Field Analysis)**» a été développée à partir de 1988 par M. Cramer [178]. Elle est en particulier utilisée pour la modélisation d'activités biologiques, pour lesquelles les méthodes classiques se révèlent parfois peu performantes. L'activité biologique d'une molécule dépend généralement de son interaction avec un récepteur donné. La modélisation de cette activité peut donc être réalisée en calculant les interactions de chaque molécule (ligand) avec ce récepteur, et en établissant une relation entre ces interactions et l'activité étudiée. Cette modélisation, qui repose sur le calcul de potentiels d'interactions moléculaires, s'effectue en plusieurs étapes.

1. **Recherche des conformations moléculaires les plus stables** : Des programmes tels que «**Concord**» ou «**Corina**» [138] permettent d'accéder à une conformation de basse énergie, à partir de laquelle sont créées plusieurs conformations bio-actives.
2. **Alignement des ligands** : Les potentiels d'interactions moléculaires dépendent des orientations des structures utilisées pour leur calcul. La comparaison des potentiels de différentes molécules nécessite donc un alignement préalable de ces structures. Il s'agit d'une étape assez délicate, en particulier lorsque les composés ont des structures diverses. Cette étape est même susceptible de limiter les possibilités d'application de la méthode. L'alignement est réalisé par rapport aux groupes fonctionnels identifiés comme potentiellement pharmacophores. Il existe plusieurs méthodes pour réaliser cet alignement. La plus simple est la méthode de **superposition par sous-structures**, qui consiste à superposer les molécules qui partagent un squelette commun, selon ce squelette. L'approche par **superposition de pharmacophores** ne nécessite pas cette supposition, mais part du fait que les pharmacophores de chaque molécule sont identifiés. Il s'agit ensuite de maximiser la superposition de ces groupements entre les molécules. Il est également possible d'aligner les molécules par rapport à leurs moments dipolaires ou à leurs champs électro-statiques.

3. **Calcul des champs électrostatiques et stériques** : Les champs électrostatiques et stériques sont ensuite évalués. Pour cela, on définit autour de chaque molécule, préalablement alignée, une grille 3D. Puis on calcule, en chaque point de la grille, l'énergie d'interaction de la molécule avec un atome sonde.
4. **Corrélation entre les champs et les activités biologiques** : Les valeurs calculées des champs peuvent alors être utilisées comme variables du modèle. Le nombre de ces variables peut être très grand (quelques milliers), lorsque la résolution de la grille est fine par rapport à la taille des molécules. Il n'est alors pas possible de corrélérer directement ces valeurs de champs avec l'activité étudiée. La modélisation est généralement effectuée par la méthode des moindres carrés partiels, et le nombre optimum de variables final déterminé par validation croisée.
5. **Une visualisation graphique** des résultats permet enfin de situer autour d'une molécule les régions pour lesquelles une substitution augmente ou diminue l'affinité envers le récepteur.

Comme nous pouvions le voir, la méthode de **contribution de groupe** est une méthode d'intuition, et sur laquelle plusieurs des méthodes de modélisation qui montraient leur efficacité l'utilise. La méthode *CoMFA*, a montré son efficacité en **QSAR**, en particulier pour la modélisation d'interactions ligand-protéine (voir le paragraphe suivant). Elle permet en effet de décrire efficacement les interactions ligand-récepteur, car les propriétés des ligands sont calculées à partir de leurs conformations bio-actives.

Problème 2.1. Les problèmes de La méthode *CoMFA* sont néanmoins multiples. On peut citer par exemple :

- Les résultats de modélisation dépendent de la conformation bio-active choisie. Or, la recherche de cette dernière est parfois difficile, en particulier lorsque la molécule est flexible : il existe alors un nombre important de conformations possibles près du minimum d'énergie.
- Il n'existe pas de règle générale pour l'alignement des molécules. Il s'agit d'un défaut majeur, car le modèle est très sensible à cet alignement.
- De plus, le temps de calcul est généralement assez long (30 à 60 min pour une vingtaine de molécules possédant de 30 à 50 atomes).

3 Le criblage virtuel

Le fait d'avoir accès à des bases de données de composés de plus en plus fournies a entraîné une nécessité d'utiliser des outils informatiques afin d'identifier les candidats potentiels, ce qui

permet de réduire les tests réels (sur des êtres vivants) en n'envoyant en laboratoire que les composés susceptibles de se lier à la cible d'intérêt.

Devenu ainsi depuis une dizaine d'année une partie intégrante du processus de recherche de nouvelles molécules bio-actives «**le Criblage Virtuel**» désigne l'action de rechercher à l'aide de programmes informatiques, grâce à de larges bases de données, des molécules virtuels pouvant selon la prédiction se lier à des cibles macromoléculaires ayant un intérêt pharmaceutique, ou posséder les propriétés souhaitées. L'intérêt donc du **Criblage Virtuel** est de permettre de découvrir des «composés nouveaux», dans le sens où certains composés ayant des structures inhabituelles par rapport aux ligands communément utilisés peuvent être mis à jour au cours du criblage. Ces composés, suffisamment différents des composés requêtes, peuvent potentiellement être considérés comme une nouvelle classe d'agents thérapeutiques.

Les Hits (c'est-à-dire, les molécules candidates) choisis sont envoyés pour être testés au niveau biologique et découvrir si de nouveaux composés actifs sont identifiés parmi eux.

Le Criblage Virtuel est habituellement divisé en deux sous-catégories : «**le Criblage virtuel basé sur la structure de la cible**» et «**le Criblage Virtuel ligand-based**» (basé sur la structure du ligand) [70].

3.1 Le criblage virtuel basé sur la structure de la cible

Le Criblage Virtuel basé sur la structure de la cible utilise la structure tridimensionnelle de la macro-molécule cible pour découvrir des petites molécules capables de s'y accrocher et prédire leur affinité de liaison, Ce que l'on désigne par le **scoring**. Ce processus, appelé le **docking** (ancrage), peut être efficace pour trouver des candidats mais reste compliqué à mettre en place pour des études impliquant des millions de ligands potentiels [54, 176].

Le processus de «**docking**» cherche à prédire l'orientation et la conformation d'un ligand dans le site de liaison d'une macro-molécule cible, puis, à partir de ceci, calcule l'affinité estimée du ligand pour la cible, justifiant ou pas de considérer ce ligand comme un hit potentiel [124].

3.2 Le criblage virtuel basé sur la structure du ligand

Lorsque la structure de la cible n'est pas connue, une autre méthode de Criblage Virtuel est appliquée afin de déterminer des candidats. Cette méthode, basée sur la structure et sur diverses caractéristiques des ligands connus de la cible, implique de cribler des bases de données de composés en utilisant ces informations comme requêtes [74, 79]. Il existe plusieurs manières de mettre en place un **Criblage Virtuel basé sur la structure du ligand**. Nous pouvons ainsi distinguer ces méthodes :

3.2.1 Les méthodes locales

Les méthodes locales sont basées sur des caractéristiques pré-définies comme étant déterminantes pour l'activité biologique. Auparavant, il est donc nécessaire de savoir quels sont les sous-ensembles de la structure globale qui impactent sur l'activité, cette information qui peut être extraire à partir d'exemples d'actifs et d'inactifs déjà connus. Pour l'extraire on peut faire :

- **La recherche via des modèles QSAR (Relation Structure-Activité) :** L'analyse **QSAR** est capable d'extraire l'information sur l'activité biologique d'une molécule à partir d'un ensemble de molécules de référence. (voir le précédent paragraphe).
- **La recherche par sous-structure :** La recherche par sous-structure renvoie tous les composés contenant la sous-structure requête. Cette méthode est basée uniquement sur la comparaison des composés en prenant en compte leur structure. C'est un cas particulier de **QSAR**, basé sur un modèle postulant que l'activité est conditionnée par la présence d'un fragment spécifique vu dans toutes les structures d'actifs connus, et absent dans les inactifs. Note : une recherche sous-structurale est souvent utilisée dans un contexte de planning synthétique : pour faire des amides, il faut chercher acides «**-COOH**» et amine «**N substitué par carbones saturés**» dans la base ici, le modèle structure-activité n'est pas conçu spécifiquement pour ce problème, mais il est appris en cours de chimie organique.
- **La recherche via des pharmacophores 3D :** On utilise lors d'une recherche via des pharmacophores **3D** un ensemble de conformations **3D** obtenues à partir de molécules actives et inactive de référence. Il est possible de déterminer à partir de cet ensemble quelles sont les parties des molécules qui interagissent avec la cible ce qui, encore une fois, relève du domaine du **QSAR**. L'arrangement dans l'espace de ces points d'interaction est utilisé comme requête dans la base de données et les molécules trouvées dans la base de données qui présentent des arrangements similaires au pharmacophore sont sélectionnées comme candidates.

3.2.2 Les méthodes globales

Les méthodes globales prennent en compte la structure moléculaire complète (car on ne sait pas où là-dedans se cachent les vrais points-clés définissant l'activité). Elles se basent sur le principe de la similitude : «**des molécules similaires ont une plus forte chance de présenter des activités similaires**» (par rapport à n'importe qu'elle paire de composés choisie aléatoirement).

Les **Hits**, dans ce contexte-là, seront alors les molécules les plus similaires à la structure d'un actif connu, le composé-requête. Il reste à définir sur quelle base cette similitude sera évaluée :

- **Recherche basée sur les graphes :** Dans cette méthode, les molécules sont représentées comme des graphes, c'est-à-dire des ensembles de sommets (les atomes) reliés entre eux

par des arêtes (les liaisons). La recherche peut ainsi se faire par sous-graphes, en utilisant des parties de la molécule comme requête ou par graphe complet. On va donc rechercher la meilleure correspondance entre les atomes et les liaisons de la requête par rapport aux atomes et aux liaisons des composés de la base de données (c'est-à-dire le sous-graphe connexe commun maximal, contenant le plus de noeuds). Cette méthode peut prendre ou non en compte les propriétés physico-chimiques des atomes et des liaisons.

- **Recherche basée sur la superposition** : Cette technique essaye de superposer une molécule sur une autre. Pour les graphes moléculaires, cette superposition, improprement qualifiée de «2D» se fait en cherchant la correspondance entre les atomes de la molécule *A* et les atomes de la molécule *B*. En 3 dimensions, elle implique de trouver la meilleure superposition entre les deux objets tridimensionnels, en se basant soit sur les distances entre atomes, soit sur une distance par exemple entre les champs entourant les atomes. Il existe de nombreuses méthodes de superposition [11, 129, 182].
- **Recherche basée sur des descripteurs** : Cette méthode considère les molécules comme un ensemble de descripteurs (habituellement chiffrés), représentant des propriétés structurales ou physico-chimiques. Ainsi, une molécule est considérée comme étant un point dans un espace multidimensionnel de descripteurs : indices topologiques, propriétés physico-chimiques calculées, histogrammes de distribution des différentes propriétés locales (comptage de fragments, de multiplets d'atomes, etc). On détermine la similarité entre deux molécules grâce à des fonctions scoring [70].

Deuxième partie

Les méthodes de calcul de l'indice de Wiener et ses extensions

Chapitre 3

L'indice de Wiener

Puisque la majorité des substances chimiques sont composées de molécules, et ces dernières possèdent **une structure**, ce qui est connu depuis la moitié du dix-neuvième siècle, l'un des principaux objectifs de la chimie organique, est d'établir une relation entre des propriétés physico-chimiques des substances et leur structure moléculaire. En générale, ces propriétés (chimiques en particulier) connues comme nous avons vu au précédent chapitre et depuis la fin du 19^{ème} siècle par **les groupes fonctionnels**, sont venues de la nature. L'évident exemple est, lorsque le composé possède la molécule $-COOH$ ça explique automatiquement l'existence de l'acide dans la substance comportant cette molécule. L'ancienne tendance de la chimie organique est d'aller plus loin, des «**méthodes de contribution des groupes**», et d'essayer de quantifier, puis de classer les molécules composant des composés plus complexes. Mais s'ai été toujours confronté par une problématique persistée dans les esprits de plusieurs scientifiques (informaticiens, chimistes, biologistes et même mathématiciens) depuis la fin des années quatre-vingts. Elle se base sur le fait qu'une structure moléculaire est une notion non numérique, par contre les propriétés physico-chimiques contenues dans les molécules sont des valeurs numériques.

Plusieurs solutions ont été proposées. Mais, la seule sur laquelle, les modélateurs ont été d'accord, est de quantifier tout d'abord l'information contenue dans la structure moléculaire, c'est-à-dire, on doit les transformer en des valeurs numériques qui les reflètent. Ce qui va les rendre plus ou moins équivalentes, et rendre aussi leur corrélation possible. Mais, jusqu'au début des années quatre-vingts, personne ne savait comment faire cette transformation, malgré les nombreuses tentatives qu'elles ont eues lieu dans ce sens. Mais il nous fallait attendre quelques chercheurs qui vont utiliser le soi-disant «**l'indice topologique**». Un **indice topologique** est un nombre réel qui est en quelque sorte calculé à partir de la molécule après l'avoir représentée en un graphe planaire, et pour lequel on croit qu'il reflète ses caractéristiques structurelles pertinentes.

Le présent chapitre étudiera l'un des plus importants et anciens indices topologiques, l'**indice de Wiener**. En donnant tout d'abord sa définition de base, ses applications sur des graphes

particuliers afin de savoir la complexité de cette définition, comme nous allons donner à travers ce chapitre quelques utilités des indices topologiques en prenant comme exemple l'indice de Wiener. Puis nous consacrons la deuxième section à nos premières contributions, qui portent sur la modification de la formule de définition de l'indice de Wiener et ses applications sur des graphes particuliers. Le troisième section dans ce chapitre va nous parler sur le type particulier des graphes le plus utilisé en théorie des graphes moléculaire, les arbres, en donnant quelques méthodes pour calculer leur indice de Wiener, afin de valoriser et voir l'importance de notre contribution. Ce chapitre terminera par une section traitant une notion très importante en graphe moléculaire en particulier et la théorie des graphes en générale, c'est **la conjecture de Wiener**.

1 La définition de base et ses applications

1.1 L'histoire de l'indice de Wiener

En 1947, Harold Wiener a publié son article [98] intitulé par «Structural Determination of Paraffin Boiling Points», dans lequel il a défini une nouvelle grandeur, qui va changer radicalement, les visions des chercheurs dans différentes disciplines, en particulier, la pharmacologie, la chimie, la physique, la télécommunication... Cette notion était l'indice de Wiener. Les files de cette histoire ont débuté entre les années 1947 et 1948, lorsque H. Wiener publiait une série des articles [96–100] où il a prouvé l'existence d'une excellente corrélation entre l'indice de Wiener et différentes propriétés physico-chimies des composés organiques, ce qui va donner les premières vues d'une nouvelle discipline. Mais, il nous fallait attendre quinze ans après que L. Steil et G. Thodos [145] entreprenaient de nouveau à donner de l'énergie aux scientifiques pour qu'ils s'approfondissent cette discipline. Parmi les scientifiques, qui vont répondre à l'appel de Steil et Thodos, on trouve H. Hosoya [81] qui va proposer en 1971 une autre définition plus applicable et plus générale de l'indice de Wiener, puis, les années 1975 et 1976 vont connaître l'émergence de deux autres scientifiques, D. H. Rouvaray et B. C. Crafford, qui ont pensé que cet indice a été mal popularisé, ce qui va les pousser à le réinventer de nouveau, pour qu'il serait largement connu chez les théoriciens et les mathématiciens chimistes, cette initiative va rapidement, dans la moitié des années 1970, donner à l'indice de Wiener une très large bonne réputation, prouvée par une large série des articles publiés sur l'indice de Wiener [44, 49].

Les débuts des années 1990, vont connaître un étrange phénomène : c'est l'apparition d'un grand nombre des indices topologiques, qui eux aussi se basent en leurs définitions sur la notion de distance et de degrés, ce qui va les rendre comme des extensions de l'indice de Wiener.

1.2 La définition de l'indice de Wiener

Dans le jargon de la théorie des graphes, la valeur numérique représentant la structure topologique d'une molécule après sa représentation en graphe est désignée par «**Descripteur Topologique**». Les descripteurs topologiques sont des invariants structurels, qui ne dépendent ni de l'étiquetage du graphe, ni de l'angle dont on voit le graphe. En d'autre terme, ils sont les valeurs calculées à partir d'un graphe projeté différemment, et qui ne se changent pas même si on change l'angle de projection. Lorsqu'un descripteur topologique corrèle une propriété (physique ou chimique) d'une molécule, il devient «**un indice moléculaire**» ou «**un indice topologique**». L'indice topologique est, donc, un descripteur topologique qui doit être capable de représenter une propriété (activité) moléculaire [163]. La représentation des propriétés (structures) moléculaires en des vecteurs numériques, rend l'existence de la molécule sur l'ordinateur une chose possible. Cette existence rend par la suite la prévision d'autres propriétés ou activités possible et d'une façon rapide. Ce processus n'est que l'une des plus importantes étapes de la découverte des nouveaux médicaments en utilisant les ordinateurs. Comme l'a été défini au

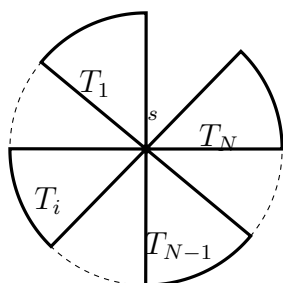


Figure 3.1 – Exemple de graphe Star-Trees

début par H. Wiener [98], l'indice de Wiener d'un graphe G , noté $W(G)$, est égale à la somme de toutes les distances entre toutes les paires de sommets dans un graphe :

$$W(G) = \sum_{\{u,v\} \subseteq V(G)} d(u,v) \quad (3.1)$$

$$= \frac{1}{2} \sum_{u \in V(G)} \sum_{v \in V(G)} (d(u,v)) \quad (3.2)$$

$$= \frac{1}{2} \sum_{u \in V(G)} w(u,G) \quad (3.3)$$

Où $w(u, G)$ nomé l'indice de Wiener d'un sommet u du graphe G est définie en [150, 151, 168] par :

$$w(u, G) = \sum_{v \in V(G)} d(u, v) \quad (3.4)$$

C'est une notion introduite depuis la fin des années 1980 en plusieurs articles, par exemple [150, 151, 168]. Et qui sera beaucoup utilisée dans cette thèse. C'est une valeur sommant les différentes distances entre le sommet u et les autres sommets du graphe.

Le prochain paragraphe connaîtra l'application directe de ces précédentes définitions sur quelques graphes, afin de toucher la difficulté de leurs manipulations directes d'une part et pour les se familier d'autre part. Les suivants travaux sont publiés dans [149], ainsi qu'en d'autres papiers scientifiques.

1.3 L'utilisation de définition de l'indice de Wiener

Nous allons utiliser dans ce paragraphe les définitions 3.1 et 3.4 pour calculer l'indice de Wiener de quelques graphes simples connus dans la littérature comme le graphe **Star-Trees** construit de N arbres connexes par un seul sommet, puis d'autres graphes apparus plus compliqués comme le graphe de **Path-trees** afin de montrer la difficulté de leurs applications dans les situations générales. Nous voulons juste signaler que tous les résultats qui seront vus dans ce paragraphe

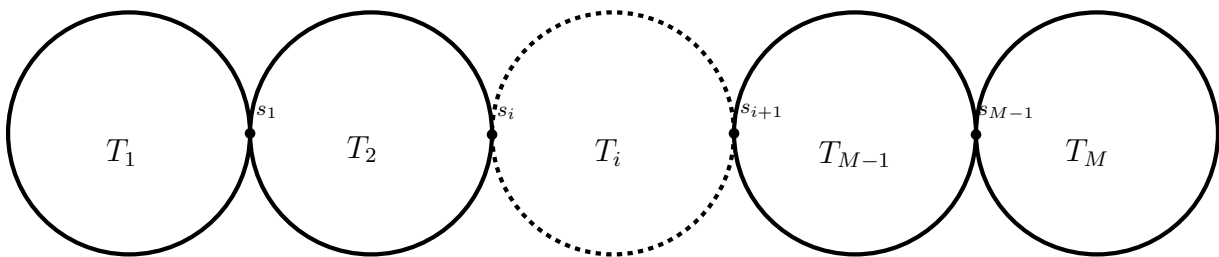


Figure 3.2 – Exemple de graphe Path-Trees

ne seront pas démontrés car, leurs démonstrations sont longs et évidentes.

1.3.1 Le graphe Star-trees

On définit un graphe Star-trees noté T_N tout graphe construit de N arbres notés T_i chaque arbre T_i est d'ordre n_i , pour $i = 1, 2, \dots, N$. Les arbres T_i sont connexes tous par un sommet s (voir la Figure 3.1).

3.1 La définition de base et ses applications

Théorème 3.1. [149][77] Soit T_N un graphe Star-trees, donc l'indice de Wiener de ce graphe est :

$$W(T_N) = \sum_{i=1}^N W(T_i) + \sum_{i=1}^{N-1} \sum_{j=i+1}^N [(n_j - 1)w(s, T_i) + (n_i - 1)w(s, T_j)] \quad (3.5)$$

La démonstration de ce théorème est facile mais un peu longue, voir [149] et [77]. Maintenant si on prend le cas où tous les arbres T_i possèdent le même ordre n (c'est-à-dire $n_i = n$ pour $i = 1, \dots, n$), on notera cet arbre par T . Le théorème 3.1 devient :

Proposition 3.1. [149] Soit T_N le graphe de Star-trees construit de N mêmes arbres T d'ordre n , donc :

$$W(T_N) = NW(T) + N(N - 1)(n - 1)w(s, T) \quad (3.6)$$

Pour plus de détail et résultats sur ce type de graphes voir [1] et [77].

1.3.2 Le graphe Path-trees

On appelle un graphe Path-trees noté T_M tout graphe formé de M arbres T_i , chacun de ces arbres d'ordre n_i concaténés les uns aux autres par un sommet s_i , pour $i = 1, 2, \dots, M - 1$, l'arbre T_i est relié à l'arbre T_{i+1} par le sommet s_i (voir la Figure 3.2).

Théorème 3.2. [149][77] Soit un graphe Path-trees noté T_M . Donc, l'indice de Wiener de ce graphe est :

$$W(T_M) = \sum_{i=1}^M W(T_i) + \sum_{i=1}^{M-1} [(n_{i+1} - 1)w(s_i, T_i) + (n_i - 1)w(s_i, T_{i+1})] + \sum_{i=1}^{M-2} \sum_{j=i+1}^M [(n_j - 1)w(s_i, T_i) + (m_i - 1)w(s_{j-1}, T_j) + (m_i - 1)(m_j - 1)d(s_i, s_{j-1})] \quad (3.7)$$

La même chose que le précédent résultat, la démonstration de la formule 3.7 demande beaucoup de faciles calculs, voir [149] et [77] pour plus de détail. Le cas particulier du theoreme 3.2 est lorsqu'on considère les arbres T_i tous de même ordre n (c'est-à-dire $n_i = n$ pour $i = 1, 2, \dots, M$).

Proposition 3.2. [149][77] Soit T_M le graphe Path-trees construit de M mêmes arbres T d'ordre n . Donc, l'indice de Wiener de ce graphe est :

$$W(T_M) = MW(T) + M(n - 1)(M - 1)w(s_1, T) + \frac{M(M - 2)(M - 1)(n - 1)^2}{6}d(s_1, s_2) \quad (3.8)$$

Pour plus de détail et résultats sur ce type de graphes voir [1] et [77].

1.4 Les utilités de l'indice de Wiener

Malgré la perte probable de l'information cachée dans les composés chimiques en raison de l'identification des molécules par des valeurs numériques, les descripteurs topologiques ont eu diverses applications dans différents domaines. Dans la dernière décennie, ils sont beaucoup utilisés dans le processus de la découverte des nouveaux médicaments, ou ce que les pharmaciens le désignent par «**Drug discovery**» (voir le chapitre II et [26][48] pour plus d'informations). Ils sont aussi appliqués dans les tests de similarité, l'isomorphisme (voir le chapitre II, et [155][161] pour plus de détails) et dans d'autres nouvelles disciplines comme par exemples le «**Graph Mining**» apparu plus récemment [43][71].

Les premières utilités de cet indice est apparu dès le début dans le papier [98], lorsque H. Wiener va le déclarer dans sa formule (la formule de Wiener pour les points d'ébullition t_B) au but de calculer les points d'ébullition des Alcanes. La formule est définie comme suit :

$$t_B = \alpha.W + \beta.p + \gamma \quad (3.9)$$

Où α , β et γ , sont des constantes expérimentales, et $p = d_G(3)$ est le nombre de polarité (le nombre de paires de sommets ayant la distance égale à trois).

Dans ses articles [96, 97, 99, 100] publiés après [98], H. Wiener a déterminé l'efficacité de $W(G)$ dans la recherche des relations structure-propriétés [97] (c'est-à-dire, la relation entre la structure de molécule et ses propriétés). Il l'a utilisé, donc, pour prédire les point d'ébullitions, les volumes molaires, les indices de réfraction, les températures d'isomérisation et les températures de vaporisation des Alcanes...

Depuis 1976, l'indice de Wiener a reconnu des différentes remarquables applications chimiques. Ces applications, qui sont en plus qu'il ait été juste des théorèmes mathématiques fondamentales, il est présenté directement en détail en plusieurs documents mono-graphes [18, 39][42] et différents papiers scientifiques [165, 174][45, 50, 123, 186]. Autrement dire, l'indice de Wiener devenait le plus fréquemment et efficace descripteur structurel employé, en le déduisant d'un graphe moléculaire. Aux années 1990, il a été démontré que l'indice de Wiener mesure l'aire de la surface des molécules, ce qui reflète donc son compacte. Et par conséquence, sa relation à la force intermoléculaire [165][115, 116] et spécialement dans le cas des hydrocarbonate où les groupes polaires sont absents. Les propriétés physiques et chimiques des substances organiques, qui peuvent être dépendent à l'aire de la surface moléculaire et/ou aux liaisons de squelette moléculaire d'atomes du carbone, ils sont souvent bien corrélés avec l'indice de Wiener. Parmi eux on trouve les températures de formation, de vaporisation et d'atomisation. Il est aussi cor-

rélié à la densité, au point d'ébullition, à la pression critique, à l'indice de réflexion, à la tension de surface, ... (voir [162] pour plus de détails).

Puisque les activités pharmacologiques des substances sont liées aux quelques ses propriétés physico-chimies (substances), il n'est pas donc étonnant, que des essais ont été faits pour utiliser, comme nous l'avons vu au chapitre II, l'indice de Wiener dans le processus de **Drug discovery** [47, 125, 126], (voir le chapitre II), et par lequel va se connaître par la suite.

2 La formule de l'indice de Wiener améliorée

A travers la section que nous venons de la voir, nous avons constaté que l'utilisation de définition de l'indice de Wiener est beaucoup plus complexe. Cette complexité s'agrandie autant que l'ordre du graphe augmente. Or, la plupart des graphes chimiques sont plus complexe (plusieurs molécules se relient l'une à l'autre), ce qui demande d'élargir la problématique du calcul de l'indice de Wiener et d'essayer, par conséquence, de donner des formules valables non seulement pour les graphes simples mais aussi pour n'importe quel graphe connexe, fini et planaire.

2.1 La formule de l'indice de Wiener en fonction du diamètre

Lorsqu'on examine les projets qu'on a été réalisés concernant le calcul de l'indice de Wiener ou ses extensions dans la récente ou ancienne littérature, on constate qu'ils traitent des graphes particuliers, par exemple, «**le graphe d'One-Heptagonal Carbon Nanocone**» [147], «**les graphes de Thorn**» [19], les graphes représentant «**le Benzenoid**» [195][61], «**la chaîne des Hexagonals**» [187], ou divers graphes.

L'idée de modifier la formule de définition de l'indice de Wiener s'impose. On s'intéresse dans cette section, comme il était déjà fait mais juste pour les graphes de diamètre 2 (voir [148][168][122]), à l'écriture de la formule de l'indice de Wiener $W(G)$ en fonction du diamètre du graphe G et des $d_G(k)$ (le nombre de paires de sommets ayant la même distance k). Commençant par les résultats concernant les graphes de diamètre 2.

Théorème 3.3. [148][168][122] *Soit G un graphe d'ordre n , de taille m et son diamètre est $D(G) = 2$. Donc, l'indice de Wiener de G peut être écrit comme suit :*

$$W(G) = n(n - 1) - m. \tag{3.10}$$

Sa démonstration est évidente.

Chapitre 3. L'indice de Wiener

Démonstration. Soit G un graphe de n sommets, et m arêtes, donc :

$$\begin{aligned}
 W(G) &= \frac{1}{2} \sum_{u \in V(G)} \sum_{v \in V(G)} (d(u, v)) \\
 &= \frac{1}{2} \sum_{u \in V(G)} \sum_{v \in V(G)} d(u, v) = \frac{1}{2} \sum_{u \in V(G)} \sum_{v \in V(G)} d(u, v) \\
 &= |E| + \sum_{u \in V(G)} (n - \deg(u) - 1) \\
 &= n(n - 1) - m.
 \end{aligned}$$

□

Définition 3.1. On appelle le graphe Maximale planaire noté \mathcal{E}_n tout graphe d'ordre n dans lequel chacun de ses sommets possède le même ordre. Autrement dit, le graphe Maximale planaire est le graphe dont ses sommets sont tous adjacents entre eux et donc de degré n ($\deg(v_i) = n$, pour $i = 1, \dots, n$) voir la Figure 3.3.

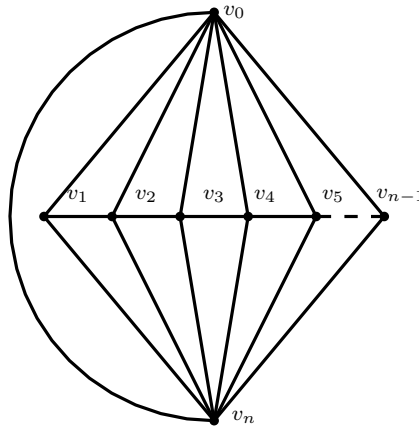


Figure 3.3 – Exemple du graphe Maximale planaire \mathcal{E}_n d'ordre n

Le résultat le plus important de ce que nous venons de voir, est celui disant que la valeur de l'indice de Wiener $W(G)$ de tout graphe fini, connexe, planaire et simple est comprise entre celle du **graphe Maximale planaire** $W(\mathcal{E}_n)$ et de **graphe Path** (une suite des arêtes concaténées les unes aux autres par un sommet v_i voir la Figure 3.13) $W(\mathcal{P}_n)$ [148][168][150], et qui a été démontrée depuis 1996 dans [104].

Théorème 3.4. [148][168][150][104] Soient G un graphe fini, simple, connexe et planaire

3.2 La formule de l'indice de Wiener améliorée

d'ordre n , \mathcal{E}_n le graphe Maximale planaire et \mathcal{P}_n le graphe Path définis avant. Donc,

$$W(\mathcal{E}_n) \leq W(G) \leq W(\mathcal{P}_n) \quad (3.11)$$

Démonstration. Pour la démonstration de ce théorème, on peut consulter les articles [104][168][148][150].

□

Nous avons modifié nous aussi la formule de définition de l'indice de Wiener, on la récrit en fonction de diamètre $D(G)$ et $d_G(k)$ (le nombre de paires de sommets ayant la même distance k) du graphe G dont on cherche son indice de Wiener. Cette nouvelle écriture, qui est la généralisation de la formule 3.10 vue précédemment, facilite le calcul de l'indice de Wiener pour la majorité des graphes en particulier ceux que nous pouvons trouver facilement leur $d_G(k)$ et qui sont finis, simples, planaires et connexes.

Théorème 3.5. [150][151][3][5][4][7] Soit G un graphe connexe, fini, non orienté, planaire et simple, d'ordre n , et de taille m . Donc,

$$\begin{aligned} W(G) &= n(n-1) - m + d_G(3) + 2d_G(4) + \dots + (D-2)d_G(D) \\ &= n(n-1) - m + \sum_{k=3}^D (k-2)d_G(k). \end{aligned}$$

Démonstration. Posant, tout d'abord :

$$\begin{aligned} W(G) &= W_1(G) \\ &= d_G(1) + 2d_G(2) + 3d_G(3) + \dots + Dd_G(D). \end{aligned}$$

Nous avons, donc :

$$\begin{aligned} \frac{n(n-1)}{2} &= d_G(1) + d_G(2) + d_G(3) + \dots + d_G(D) \\ \frac{n(n-1)}{2} &= m + d_G(2) + d_G(3) + \dots + d_G(D) \\ d_G(2) &= \frac{n(n-1)}{2} - m - d_G(3) - \dots - d_G(D). \end{aligned}$$

On remplace dans la première équation et on déduit notre formule.

□

Lorsqu'on examine la formule de Wiener pour calculer les points d'ébullition 3.9, que nous avons vu à la précédente section, on trouve, que parmi ses parties la valeur P (nombre de

polarisation) qui n'est que $d_G(3)$. Donc si on veut la récrire en remplaçant l'indice de Wiener par notre formule, on trouve si $\alpha \neq 0$:

$$\begin{aligned} t_B &= \alpha W + \beta P + \gamma \\ &= \alpha[n(n-1) - m + d_G(3) + 2d_G(4) + \dots + (D-2)d_G(D)] + \beta d_G(3) + \gamma \\ &= \alpha[n(n-1) - m + \left(\frac{\beta}{\alpha} + 1\right)d_G(3) + 2d_G(4) + \dots + (D-2)d_G(D)] + \gamma \end{aligned}$$

Dans le paragraphe suivant, nous allons utiliser la formule 3.5 pour calculer l'indice de Wiener de quelques graphes connus dans la littérature. Cette application va nous confirmer l'utilité de cette réécriture pour quelques catégories de graphes.

2.2 L'utilisation de la formule avec des graphes à faible diamètre

Dans ce paragraphe, nous allons citer les applications de la formule 3.5 sur des classes prédéfinies de graphes qui sont classés selon leurs diamètres. Nous commencerons cette application par les graphes de diamètre 2, que nous allons les décrire au début du prochain paragraphe.

2.2.1 Les graphes de diamètre 2

Définition 3.2. On appelle un graphe Fan d'ordre n noté \mathcal{F}_n , tout graphe de taille $m = 2n - 3$ et dans lequel on a un sommet de v_0 de degré $deg(v_0) = n - 1$ adjacent à tous les autres sommets. Les autres sommets du graphe forment un graphe path d'ordre $n - 1$ dont ses sommets sont tous de degré $deg(v_i) = 3$, pour $i \neq 0, 1, n$ sauf les sommets v_1 et v_n que leur degrés égaux à 2 $deg(v_j) = 2$ pour $j = 1, n$ (voir la Figure 3.4).

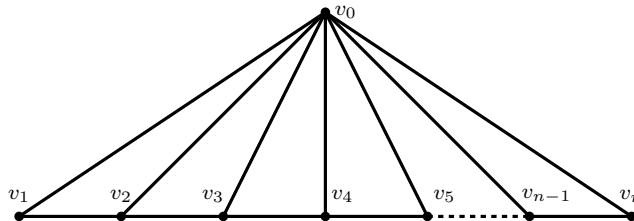


Figure 3.4 – Exemple de graphe Fan \mathcal{F}_n d'ordre n

Définition 3.3. On appelle un graphe Wheel d'ordre n noté \mathcal{W}_n , tout graphe de taille $m = 2n - 2$ et dans lequel on a un sommet de v_0 de degré $deg(v_0) = n - 1$ adjacent à tous les autres sommets. Les autres sommets du graphe forment un graphe Cycle d'ordre $n - 1$ dont ses sommets sont tous de degré $deg(v_i) = 3$, pour $i \neq 0$ (voir la Figure 3.5).

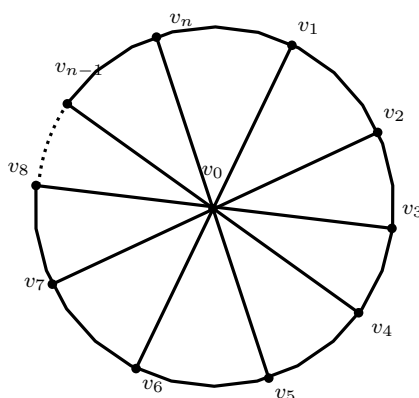


Figure 3.5 – Exemple du graphe Wheel \mathcal{W}_n d'ordre n

Définition 3.4. Le graphe Butterfly d'ordre n noté \mathcal{B}_n est un graphe de taille $m = 2n - 4$ construit de $n - 3$ sommets de degré $\deg(v_i) = 2$, pour $i \neq 2, 3, 4$ et des trois autres sommets sont de degré différents à 2. C'est un graphe qui schématise un aéroport connectant trois villes ou pays possédant tous plusieurs autres aéroports (voir la Figure 3.6).

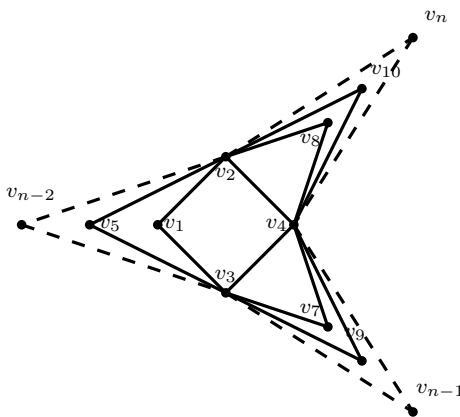


Figure 3.6 – Exemple du graphe Butterfly \mathcal{B}_n d'ordre n

Définition 3.5. On appelle un graphe Crystal d'ordre n noté \mathcal{C}_n tout graphe Maximale planaire \mathcal{E}_n d'ordre n en élevant l'arête reliant les deux sommets v_0 et v_n . Le graphe Crystal est un graphe de taille $m = 3n - 7$ (voir la Figure 3.7).

Après ces brèves définitions de graphes de diamètre 2, nous appliquons la formule 3.5 sur ces graphes.

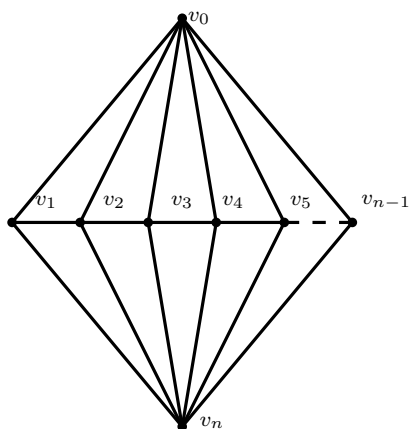


Figure 3.7 – Exemple du graphe Crystal \mathcal{C}_n d'ordre n

Théorème 3.6. [150][151] Soient \mathcal{F}_n , \mathcal{W}_n , \mathcal{E}_n , \mathcal{B}_n , et \mathcal{C}_n les graphes Fan, Wheel, Maximale planaire, Butterfly et Crystal respectivement comme ils sont définis avant. Donc leurs indices de Wiener sont :

G_n	$W(G_n)$	n
\mathcal{F}_n	$n^2 - 3n + 3$	$n \geq 3$
\mathcal{W}_n	$n^2 - 3n + 2$	$n \geq 5$
\mathcal{E}_n	$n^2 - 4n + 6$	$n \geq 3$
\mathcal{B}_n	$n^2 - 3n + 4$	$n \geq 4$
\mathcal{C}_n	$n^2 - 4n + 7$	$n \geq 5$

Démonstration. En appliquant juste soit le théorème 3.5 soit le théorème 3.3 car ces graphes sont de diamètre 2. □

2.2.2 Les graphes de diamètre 3

Nous discuterons dans la classe des graphes de diamètre 3, un exemple de graphe que nous les appelons le graphe Double-Crystal. Le graphe Double-Crystal d'ordre toujours pair n noté \mathcal{DC}_n de taille $m = 3(n - 4)$ est un graphe construit de deux graphes Crystal $\mathcal{C}_{\frac{n}{2}}$ d'ordre $\frac{n}{2}$ concaténés par deux arêtes, une reliant les deux sommets en haut des deux graphes Crystal et l'autre reliant les deux sommets en bas de ces deux graphes, ces quatre sommets reliés v_0 , $v_{\frac{n-2}{2}}$, $v_{\frac{n}{2}}$ et v_{n-1} sont tous de degré $\frac{n-2}{2}$. Les quatre sommets d'extrémité des deux graphes Crystal v_1 , $v_{\frac{n}{2}-2}$, $v_{\frac{n}{2}+1}$ et v_{n-2} sont tous de degré 3 et les autres sommets restant du graphe sont tous de degré 4 (voir la Figure 3.8). Maintenant nous analysons ce graphe pour extraire

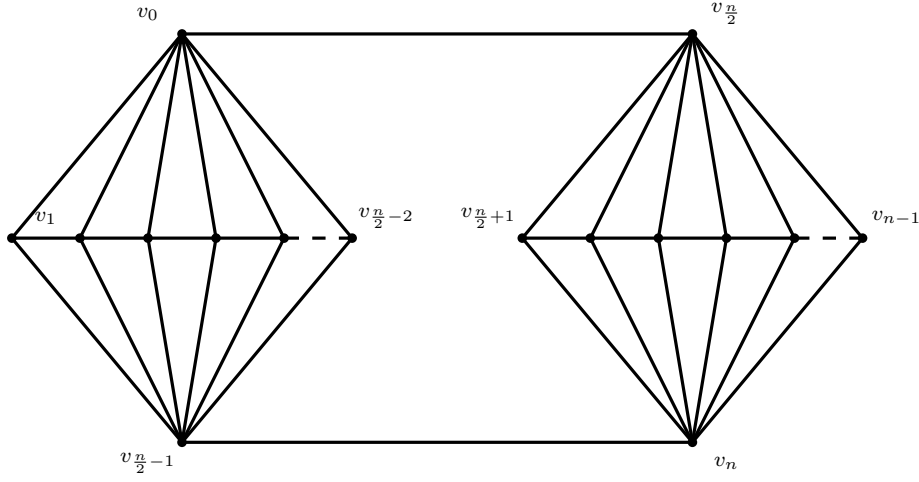


Figure 3.8 – Exemple de graphe Double-Crystal \mathcal{DC}_n d'ordre n

quelques moyens qui seront utile pour l'utilisation de la formule 3.5 avec ce graphe. Le lemme suivant montre ces moyens.

Lemme 3.1. [5][4] Soit \mathcal{DC}_n un graphe Double-Crystal d'ordre n comme nous l'avons défini précédemment. Donc,

– Pour $i = 0, 1, 2, \dots, n-1$, on a :

$$w(v_i, \mathcal{DC}_n) = \begin{cases} \frac{3}{2}n, & \text{si } i = 0, \frac{n}{2} - 1, \frac{n}{2}, n - 1 \\ \frac{5n-14}{2}, & \text{si } i = 1, \frac{n}{2} - 2, \frac{n}{2} + 1, n - 2 \\ \frac{5n-16}{2}, & \text{sinon} \end{cases} \quad (3.12)$$

– Pour $k = 1, 2, 3 = D(\mathcal{DC}_n)$, on a :

$$d_{\mathcal{DC}_n}(k) = \begin{cases} 3(n-4), & \text{si } k = 1 \\ \frac{n^2-6n+24}{4}, & \text{si } k = 2 \\ \frac{n^2-8n+24}{4}, & \text{si } k = 3 \end{cases} \quad (3.13)$$

Démonstration. Sont des simples calculs à la main. □

Théorème 3.7. [5][4] Soit \mathcal{DC}_n le graphe Double-Crystal d'ordre n , défini précédemment. Donc, l'indice de Wiener de ce graphe est :

$$W(\mathcal{DC}_n) = \frac{5}{4}n^2 - 6n + 18, \text{ pour } (n \geq 10) \quad (3.14)$$

Démonstration. En appliquant le théorème 3.5 et le Lemme 3.1. □

2.2.3 Les graphes de diamètre 4

Dans la catégorie des graphes de diamètre 4 nous traiterons deux exemples de graphes, le graphe Double-Wheel \mathcal{DW}_n et le graphe Sunflower \mathcal{FL}_n . On commence par le graphe Sunflower définie comme suit :

Définition 3.6. Le graphe Sunflower d'ordre toujours impair n et de taille $m = 2(n - 1)$ noté \mathcal{FL}_n est le graphe dont le degré de son sommet de centre v_0 est $deg(v_0) = \frac{n-1}{2}$, le degré des sommets des indices impairs v_1, v_3, \dots, v_{n-1} est $deg(v_{2i+1}) = 5$, et le degré des sommets v_2, v_4, \dots, v_n est $deg(v_{2i}) = 2$ (voir 3.9).

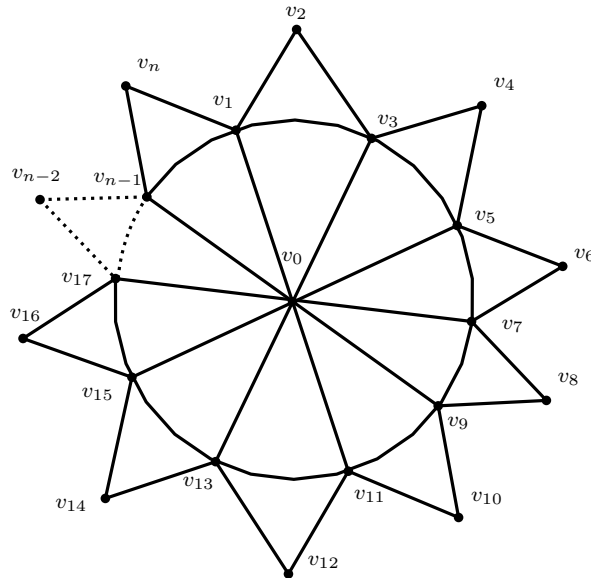


Figure 3.9 – Exemple de graphe Sunflower \mathcal{FL}_n d'ordre n

Comme avant l'utilisation du théorème 3.5 nécessite les informations suivantes :

Lemme 3.2. [5][150][151] Soit \mathcal{FL}_n le graphe Sunflower d'ordre $n \geq 11$, et de taille $m = 2(n - 1)$, défini avant. Donc,

$$w(v_i, \mathcal{FL}_n) = \begin{cases} \frac{7}{2}n - \frac{35}{2}, & \text{si } i \text{ est pair, et } i \neq 0 \\ \frac{3}{2}n - \frac{3}{2}, & \text{si } i = 0 \\ \frac{5}{2}n - \frac{23}{2}, & \text{si } i \text{ est impair et } i \neq 0 \end{cases} \quad (3.15)$$

et :

3.2 La formule de l'indice de Wiener améliorée

$d_1(\mathcal{FL}_n)$	$2n - 2$
$d_2(\mathcal{FL}_n)$	$\frac{1}{8}n^2 + n - \frac{9}{8}$
$d_3(\mathcal{FL}_n)$	$\frac{1}{4}n^2 - 2n + \frac{7}{4}$
$d_4(\mathcal{FL}_n)$	$\frac{1}{8}n^2 - \frac{3}{2}n + \frac{11}{8}$

Démonstration. Sont des simples calculs à la main. □

Théorème 3.8. [5][150][151] Soit \mathcal{FL}_n le graphe de Sunflower, défini précédemment. Donc, l'indice de Wiener de ce graphe est :

$$W(\mathcal{FL}_n) = \frac{3}{2}n^2 - 8n + \frac{13}{2} \quad (3.16)$$

Démonstration. Pour la démonstration, c'est exactement comme les précédents théorèmes, on applique la formule du théorème 3.5, avec le résultat du Lemme 3.2. □

On passe maintenant au graphe Double-Wheel que nous le définissons comme suit :

Définition 3.7. On appelle un graphe Double-Wheel d'ordre n et de taille $m = 4n - 5$ noté \mathcal{DW}_n tout graphe construit de deux graphes Wheel $\mathcal{W}_{\frac{n-2}{2}}$ d'ordre $\frac{n-2}{2}$. Les deux sommets centres des graphes Wheel construits \mathcal{DW}_n sont de degré $\frac{n-2}{2}$, dans ce graphe on trouve aussi deux autres sommets de degré 5 et les autres sont tous de degré 3 (voir la Figure 3.10).

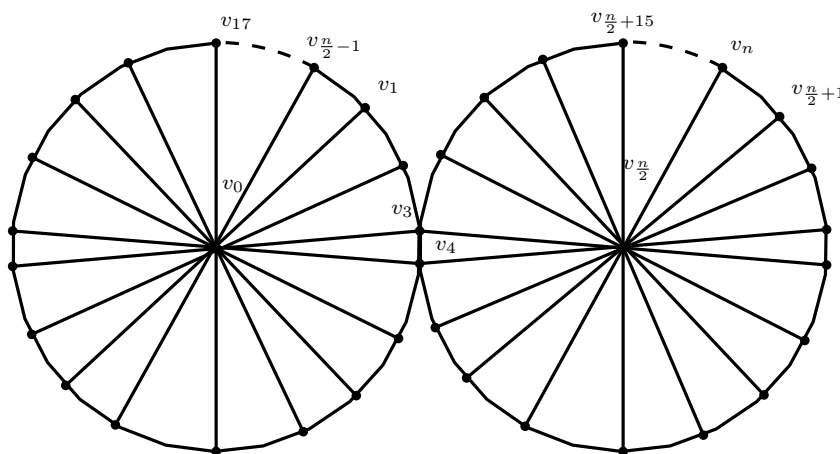


Figure 3.10 – Exemple de graphe Double-Wheel \mathcal{DW}_n d'ordre n

Lemme 3.3. [4] Soit \mathcal{DW}_n le graphe Double-Wheel d'ordre $n \geq 10$ et de taille m , défini avant. Donc,

– Pour $i = 0, 1, \dots, n - 1$, on a :

$$w(v_i, \mathcal{DW}_n) = \begin{cases} 2n - 6, & \text{si } i = 0, \frac{n}{2} + 1 \\ 2n - 7, & \text{si } i = 1, 2 \\ \frac{5n-16}{2}, & \text{si } i = 3, \frac{n}{2}, \frac{n}{2} + 2, n - 1 \\ 3n - 10, & \text{sinon} \end{cases} \quad (3.17)$$

– Pour $k = 1, \dots, 4 = D(\mathcal{DW}_n)$, on a :

$$d_{\mathcal{DW}_n}(k) = \begin{cases} 2n - 1, & \text{si } k = 1 \\ \frac{n^2-6n+28}{4}, & \text{si } k = 2 \\ 3n - 22, & \text{si } k = 3 \\ (\frac{n-8}{2})^2, & \text{si } k = 4 \end{cases} \quad (3.18)$$

Démonstration. Sont des simples calculs à la main. □

Théorème 3.9. [4] Soit \mathcal{DW}_n le graphe Double-Wheel, d'ordre n et de taille m , défini avant. Donc, l'indice de Wiener de ce graphe est :

$$W(\mathcal{DW}_n) = \frac{3}{2}n^2 - 8n + 11, \text{ pour } (n \geq 10) \quad (3.19)$$

Démonstration. Pour la démonstration, on applique la formule du théorème 3.5, avec le résultat du Lemme 3.3. □

2.2.4 Quelques graphes classique de diamètre supérieur à 4

Les graphes de diamètre supérieur à 4 sont nombreux, et nous nous limitons à quelques exemples comme, le graphe Path noté \mathcal{P}_n , Cycle noté \mathcal{O}_n , et Grid noté \mathcal{G}_n qui sont plus connus et plus faciles aux traitements. Mais avant de calculer leurs indices de Wiener nous allons les décrire.

Définition 3.8. On appelle un graphe Cycle d'ordre n et de taille $m = n$ noté \mathcal{O}_n tout graphe Path dont ses extrémités coïncident. Les sommets d'un graphe Cycle sont tous de degré $\deg(v_i = 2)$, pour $i = 1, 2, \dots, n$ (voir la Figure 3.11).

Le calcul de l'indice de Wiener du graphe Cycle demande quelques données qu'on les trouve dans le lemme suivant :

Lemme 3.4. [3][150][151][7][6] Soit \mathcal{O}_n un graphe Cycle d'ordre $n \geq 2$ et de taille $m = n$,

3.2 La formule de l'indice de Wiener améliorée

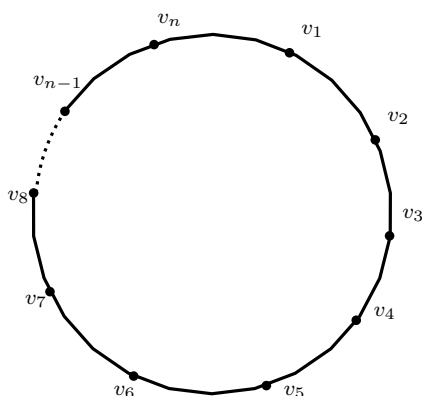


Figure 3.11 – Exemple de graphe Cycle \mathcal{O}_n d'ordre n

comme on l'a décrit avant. Donc,

$$w(v_i, \mathcal{O}_n) = \begin{cases} \frac{1}{4}n^2, & \text{si } n \text{ pair} \\ \frac{1}{4}n^2 - \frac{1}{4}, & \text{si } n \text{ impair} \end{cases} \quad (3.20)$$

et :

$$d_{\mathcal{O}_n}(i) = \begin{cases} n, & \text{si } n \text{ est pair, et } 1 \leq i < \frac{n}{2} \\ \frac{n}{2}, & \text{si } n \text{ pair et } i = \frac{n}{2} \\ n, & \text{si } n \text{ est impair, et } 1 \leq i \leq \frac{n-1}{2} \end{cases} \quad (3.21)$$

Démonstration. Sont des simples calculs à la main. □

Théorème 3.10. [3][150][151][7][6] Soit \mathcal{O}_n un graphe Cycle d'ordre $n \geq 2$ et de taille de m , comme on a décrit. Donc, l'indice de Wiener de \mathcal{O}_n est :

$$W(\mathcal{O}_n) = \begin{cases} \frac{1}{8}n^3, & \text{si } n \text{ est pair} \\ \frac{1}{8}n^3 - \frac{1}{8}n, & \text{si } n \text{ est impair} \end{cases} \quad (3.22)$$

Démonstration. Pour la démonstration, c'est exactement comme les précédents théorèmes, on applique la formule du théorème 3.5, avec le résultat du Lemme 3.4. □

Définition 3.9. Le graphe Grid d'ordre n et de taille $m = \frac{3}{2}n - 2$ noté \mathcal{G}_n est le graphe construit de deux graphes Path dont leurs sommets sont reliés un à un par des arêtes. Les sommets de ce graphe sont tous de degré 3 sauf les quatres sommets d'extrémités où leur degré est de 2 (voir la Figure 3.12).

Le lemme suivant est nécessaire pour le calcul de l'indice de Wiener du graphe Grid.

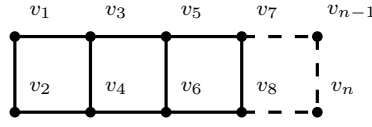


Figure 3.12 – Exemple de graphe Grid \mathcal{G}_n d'ordre n

Lemme 3.5. [3][2][4][151] Soit \mathcal{G}_n un graphe grid d'ordre $n \geq 2$ et de taille $m = \frac{3}{2}n - 2$ comme on a défini avant, donc :

– Pour $i = 1, 2, \dots, \frac{n}{2}$, on a :

$$w(u_{2i-1}, \mathcal{G}_n) = w(u_{2i}, \mathcal{G}_n) \tag{3.23}$$

$$= \frac{(n - 2i)^2}{4} + \frac{(2i - 2)^2}{4} + n - 1. \tag{3.24}$$

– Pour $k = 1, 2, \dots, \frac{n}{2} = D(\mathcal{G}_n)$, on a :

$$d_{\mathcal{G}_n}(k) = \begin{cases} \frac{3}{2}n - 2, & \text{si } k = 1 \\ 2n - 4k + 2, & \text{si } k = 2, \dots, D(\mathcal{G}_n) \end{cases} \tag{3.25}$$

Démonstration. Sont des simples calculs à la main. □

Théorème 3.11. [3][2][4][151] Soit \mathcal{G}_n un graphe grid d'ordre n et de taille m , comme on a défini. Donc, l'indice de Wiener de ce graphe est :

$$W(\mathcal{G}_n) = \frac{n^3}{12} + \frac{n^2}{4} - \frac{n}{3} \text{ pour } (n \geq 4) \tag{3.26}$$

Démonstration. L'application du théorème 3.5 et le lemme 3.5 suffisent pour la démonstration de ce théorème. □

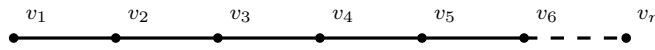


Figure 3.13 – Exemple de graphe Path \mathcal{P}_n d'ordre n

Lemme 3.6. [3][7] Soit \mathcal{P}_n un graphe Path d'ordre $n \geq 2$ et de taille $m = n - 1$ (voir la Figure 3.13). Donc,

– Pour $i = 1, 2, \dots, n$, on a :

$$w(v_i, \mathcal{P}_n) = \frac{n^2 + n}{2} - i(n + 1) + i^2 \tag{3.27}$$

– Pour $k = 1, 2, \dots, \frac{n}{2} = D(\mathcal{P}_n)$, on a :

$$d_{\mathcal{P}_n}(k) = n - k \quad (3.28)$$

Démonstration. Sont des simples calculs à la main. □

Théorème 3.12. [3][7] Soit \mathcal{P}_n un graphe Path d'ordre $n \geq 2$ et de taille m , (voir la Figure 3.13). Donc, l'indice de Wiener de ce graphe est :

$$W(\mathcal{P}_n) = \frac{n(n^2 - 1)}{6} \text{ pour } (n \geq 2) \quad (3.29)$$

Démonstration. On applique Pour la démonstration de ce théorème le théorème 3.5 et le lemme 3.6 □

2.3 L'utilisation de la formule avec les graphes composés

L'une des catégories des graphes sur lesquelles on a travaillé est celle des graphes composés de deux graphes. Dans la littérature on trouve plusieurs cas de la composition de deux graphes : le produit, le produit des arêtes, le produit des sommets, ... pour en savoir plus voir KKKK. Le cas le plus étudié dans la littérature est celui du produit des sommets ou ce qu'on peut désigner par «le produit corona». Dans ce paragraphe nous traiterons ce type de composition à travers deux exemples : Le produit Corona de deux graphes Cycle et Le produit Corona d'un graphe Cycle et Path.

2.3.1 Le produit Corona de deux graphes Cycle

Définition 3.10. Soient deux graphes Cycle des ordres respectivement n_1 et n_2 noté C_{n_1} et C_{n_2} . On appelle le produit corona des graphes C_{n_1} et C_{n_2} noté $C_{n_1} \circ C_{n_2}$ le graphe obtenu en prenant n_1 copies du graphe C_{n_2} , puis on connecte chaque $i^{\text{ème}}$ sommet du graphe C_{n_1} par une arête à chaque sommet du $j^{\text{ème}}$ copie de C_{n_2} pour $1 \leq i, j \leq n_1$. Les degrés des sommets appartenant au graphe Cycle C_{n_1} deviennent égale à n_2 et les autres sommets sont tous de degré 3 (voir la Figure 3.14) [32, 95].

Lemme 3.7. [7][6] Soit $C_{n_1} \circ C_{n_2}$ un graphe résultant du produit Corona de deux graphes Cycle d'ordre $n = n_1 + n_1 n_2$, de taille $d_{C_{n_1} \circ C_{n_2}}(1) = m = n_1 + 2n_1 n_2$ et de $d_{C_{n_1} \circ C_{n_2}}(2) = n_1 + 2n_1 n_2 + n_1 \frac{(n_2 - 3)n_2}{2}$, comme on a défini avant. Donc :

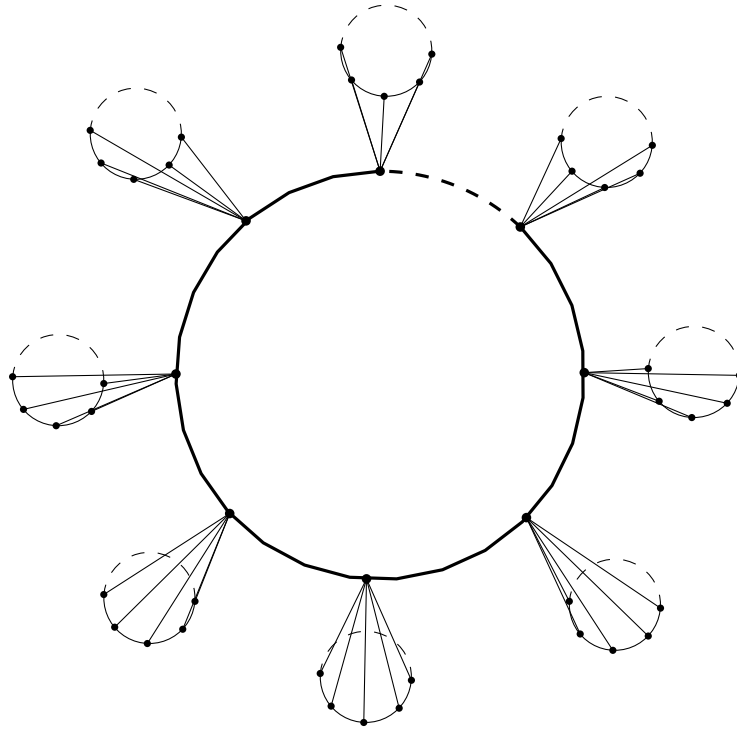


Figure 3.14 – Exemple du produit Corona de deux graphes Cycle $C_{n_1} \circ C_{n_2}$

– si n_1 est pair

$$d_{C_{n_1} \circ C_{n_2}}(k) = \begin{cases} n_1 + n_1 n_2^2 + 2n_1 n_2, & 3 \leq k \leq D(C_{n_1} \circ C_{n_2}) - 3; \\ \frac{n_1}{2} + 2n_1 n_2 + n_1 n_2^2, & k = D(C_{n_1} \circ C_{n_2}) - 2; \\ n_1 n_2^2 + n_1 n_2, & k = D(C_{n_1} \circ C_{n_2}) - 1; \\ \frac{n_1 n_2}{2}, & k = D(C_{n_1} \circ C_{n_2}). \end{cases} \quad (3.30)$$

– si n_1 est impair

$$d_{C_{n_1} \circ C_{n_2}}(k) = \begin{cases} n_1 + n_1 n_2^2 + 2n_1 n_2, & 3 \leq k \leq D(C_{n_1} \circ C_{n_2}) - 2; \\ n_1 n_2^2 + 2n_1 n_2, & k = D(C_{n_1} \circ C_{n_2}) - 1; \\ n_1 n_2^2, & k = D(C_{n_1} \circ C_{n_2}). \end{cases} \quad (3.31)$$

$$D(C_{n_1} \circ C_{n_2}) = \begin{cases} \frac{n_1}{2} + 2, & \text{si } n_1 \text{ est pair;} \\ \frac{n_1-1}{2} + 2, & \text{si } n_1 \text{ est impair;} \end{cases} \quad (3.32)$$

Démonstration. Sont des simples calculs à la main. □

Théorème 3.13. [7][6] Soit $C_{n_1} \circ C_{n_2}$ le graphe résultant du produit Corona de deux graphes Cycle d'ordre $n = n_1 + n_1 n_2$ et de taille m . Donc, l'indice de Wiener de ce graphe est :

$$W(C_{n_1} \circ C_{n_2}) = \begin{cases} \frac{n_1^3(n_2+1)^2}{8} + n_1^2 n_2 + n_1^2 n_2^2 - 2n_1 n_2, & \text{si } n_1 \text{ est pair;} \\ \frac{n_1(n_2+1)^2(n_1-1)}{8} + n_1^2 n_2 + n_1^2 n_2^2 - 2n_1 n_2, & \text{si } n_1 \text{ est impair.} \end{cases} \quad (3.33)$$

Démonstration. Puisque les deux cas (pair et impair) sont similaires, on va démontrer juste le cas où n_1 est pair, en appliquant le théorème 3.5 et le lemme 3.7.

$$\begin{aligned} W(C_{n_1} \circ C_{n_2}) &= n(n-1) - m + \sum_{k=3}^{D(C_{n_1} \circ C_{n_2})} d_{C_{n_1} \circ C_{n_2}}(k)(k-2) \\ &= (n_1 + n_1 n_2^2 + 2n_1 n_2) \sum_{k=3}^{D(C_{n_1} \circ C_{n_2})-3} d_{C_{n_1} \circ C_{n_2}}(k)(k-2) - m + n(n-1) + \\ &\quad (D(C_{n_1} \circ C_{n_2}) - 4)(n_1 n_2^2 + \frac{n_1}{2} + 2n_1 n_2) + \frac{n_1 n_2^2}{2}(D(C_{n_1} \circ C_{n_2}) - 2) + \\ &\quad (n_1 n_2 + n_1 n_2^2)(D(C_{n_1} \circ C_{n_2}) - 2) \\ &= (\frac{n_1}{2} - 2)(\frac{n_1}{2} + 2n_1 n_2 + n_1 n_2^2) - (n_1 + 2n_1 n_2) + \frac{n_1}{2}(\frac{n_1 n_2^2}{2}) + (n_1 + \\ &\quad n_1 n_2)(n_1 + n_1 n_2 - 1) + \frac{n_1^2 - 10n_1 + 24}{8}(n_1 n_2^2 + n_1 + 2n_1 n_2) \\ &\quad + (\frac{n_1}{2} - 1)(n_1 n_2 + n_1 n_2^2) \\ &= \frac{1}{8} n_1^3 (n_2 + 1)^2 + n_1^2 (n_2 + n_2^2) + n_1 (-2n_2). \end{aligned}$$

□

2.3.2 Le produit Corona des graphes Cycle et Path

Définition 3.11. Soient un graphe cycle noté C_{n_1} d'ordre n_1 et un graphe Path noté P_{n_2} d'ordre n_2 . On définit le graphe résultant du produit Corona de C_{n_1} et P_{n_2} noté $C_{n_1} \circ P_{n_2}$ le graphe obtenu en prenant n_1 copies du graphe P_{n_2} et pour chaque i , on connecte le $i^{\text{ème}}$ sommet de C_{n_1} et chaque sommet de la $i^{\text{ème}}$ copie de P_{n_2} par une arête, pour $1 \leq i \leq n_1$. Les degrés des sommets du graphe Cycle est n_2 et les autres sommets sont soit de degré 2 soit de degré 3 (Voir la Figure 3.15).

Le calcul de l'indice de Wiener du produit corona des graphes Cycle et Path, demande les données suivantes :

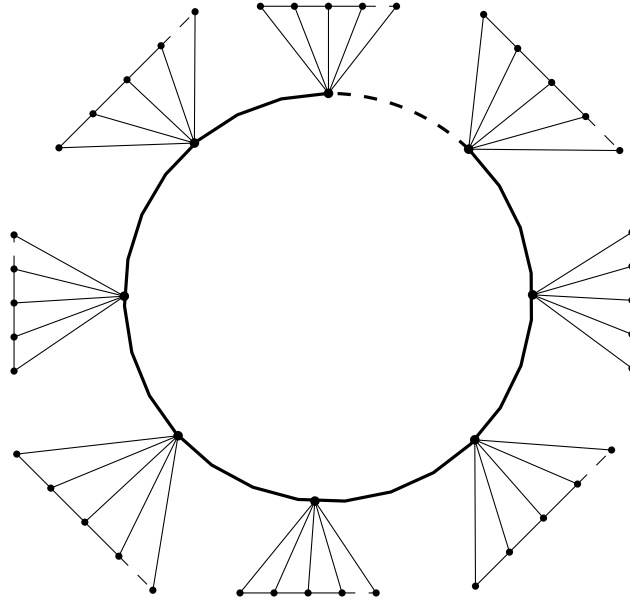


Figure 3.15 – Exemple du produit Corona des graphe Cycle et Path $C_{n_1} \circ P_{n_2}$

Lemme 3.8. [7][6] Soient $C_{n_1} \circ P_{n_2}$ le graphe résultant du produit Corona des graphe Cycle et Path d'ordre $n = n_1 + n_1n_2$, de taille $d_{C_{n_1} \circ P_{n_2}}(1) = m = 2n_1n_2$ et de $d_{C_{n_1} \circ P_{n_2}}(2) = n_1 + 2n_1n_2 + n_1 \frac{(n_2-3)n_2}{2}$, comme est défini avant. Donc,

– Si n_1 est pair

$$d_{C_{n_1} \circ P_{n_2}}(k) = \begin{cases} n_1 + n_1n_2^2 + 2n_1n_2, & 3 \leq k \leq D(C_{n_1} \circ P_{n_2}) - 3; \\ \frac{n_1}{2} + 2n_1n_2 + n_1n_2^2, & k = D(C_{n_1} \circ P_{n_2}) - 2; \\ n_1n_2^2 + n_1n_2, & k = D(C_{n_1} \circ P_{n_2}) - 1; \\ \frac{n_1n_2}{2}, & k = D(C_{n_1} \circ P_{n_2}). \end{cases} \quad (3.34)$$

– Si n_1 est impair

$$d_{C_{n_1} \circ P_{n_2}}(k) = \begin{cases} n_1 + n_1n_2^2 + 2n_1n_2, & 3 \leq k \leq D(C_{n_1} \circ P_{n_2}) - 2; \\ n_1n_2^2 + 2n_1n_2, & k = D(C_{n_1} \circ P_{n_2}) - 1; \\ n_1n_2^2, & k = D(C_{n_1} \circ P_{n_2}). \end{cases} \quad (3.35)$$

$$D(C_{n_1} \circ P_{n_2}) = \begin{cases} \frac{n_1}{2} + 2, & \text{si } n_1 \text{ est pair;} \\ \frac{n_1-1}{2} + 2, & \text{si } n_1 \text{ est impair;} \end{cases} \quad (3.36)$$

Démonstration. Sont des simples calculs. □

Théorème 3.14. [7][6] Soit $C_{n_1} \circ P_{n_2}$ le graphe résultant du produit Corona des graphe Cycle

et Path d'ordre $n = n_1 + n_2 n_1$ et de taille m , comme est défini avant. Donc, l'indice de Wiener de ce graphe est :

$$W(C_{n_1} \circ P_{n_2}) = \begin{cases} \frac{n_1^3(n_2+1)^2}{8} + n_1^2 n_2 + n_1^2 n_2^2 - n_1(2n_2 + 1), & \text{si } n_1 \text{ est pair;} \\ \frac{n_1(n_2+1)^2(n_1^2-1)}{8} + n_1^2 n_2 + n_1^2 n_2^2 - n_1(2n_2 + 1), & \text{si } n_1 \text{ est impair.} \end{cases} \quad (3.37)$$

Démonstration. La démonstration de ce théorème est similaire à celle 2.3.1 du théorème 3.13. Pour adapter la démonstration 2.3.1 au théorème 3.14 on remplace juste le nombre des arêtes de la première par $2n_1 n_2$. □

2.4 L'indice de Wiener pour le graphe Cobweb

Dans ce paragraphe nous aborderons un type spéciale des graphes de diamètre supérieur à 4. Le graphe Cobweb qu'à traverser lui nous monterons quelques difficultés qui peuvent rencontrer le théorème 3.5 et qui se voit au recherche de $d_G(k)$ du graphe en question. Pour ce paraît graphe, on trouve quelques difficultés à la recherche de $d_G(k)$, ce qui exige l'utilisation de la définition d'origine.

Définition 3.12. On appelle un graphe spider's web (Cobweb) d'ordre $n = Lp$ et de taille $m = Lp + (L - 1)p$ noté \mathcal{WB}_n tout graphe construit de L graphes Cycle d'ordre p , et chaque graphe Cycle se situe au niveau l . Tous les sommets de tous les graphes Cycle composant le graphe Cobweb sont reliés par un graphe Path d'ordre L . Tous les sommets du graphe Cobweb sont de degré 4 sauf ceux du premier niveau et du dernier niveau qui sont de degré 3 (voir la Figure 3.16).

Le calcul de l'indice de Wiener de \mathcal{WB}_n demande sa définition car la recherche de $d_{\mathcal{WB}_n}(k)$ est difficile, mais le calcul de l'indice de Wiener du sommet $u_{i,l}$ $w(u_{i,l}, \mathcal{WB}_n, u)$ où $u_{i,l}$ est le sommet d'indice i qui se trouve au niveau l dans \mathcal{WB}_n .

Lemme 3.9. [3][2][151] Soit \mathcal{WB}_n un graphe Cobweb d'ordre n et de taille m , comme nous l'avons défini avant. Donc, l'indice de Wiener du sommet d'indice i , qui se trouve au niveau l , c'est-à-dire $w(u_{i,l})$ est :

– Pour $i = 1, \dots, p$, et $l = 1, \dots, L$

$$w(u_{i,l}, \mathcal{WB}_n) = \begin{cases} pl(l-1) + \frac{L}{4}((p+1)^2 + 2Lp - 4lp - 2), & \text{si } p \text{ impair.} \\ pl(l-1) + \frac{L}{4}((p+1)^2 + 2Lp - 4lp - 1), & \text{si } p \text{ est pair} \end{cases} \quad (3.38)$$

Démonstration. Sont des simples calculs à la main. □

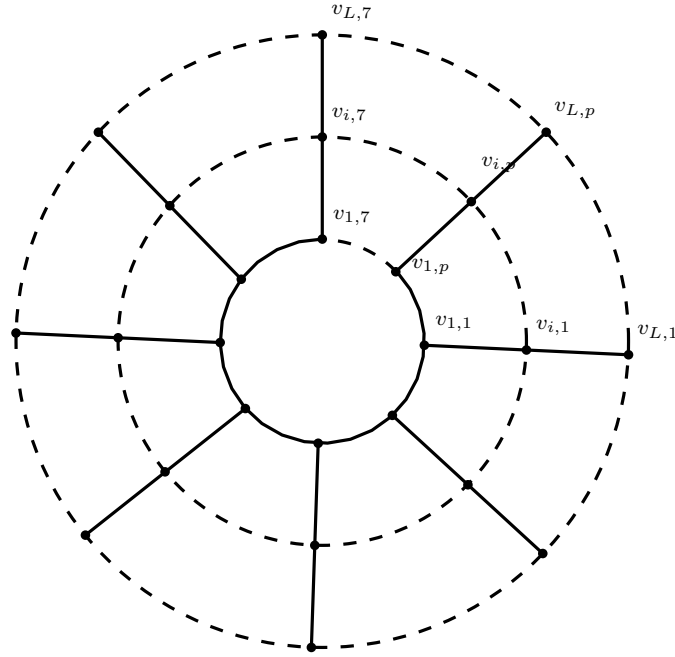


Figure 3.16 – Exemple du graphe Cobweb \mathcal{WB}_n d'ordre n formé de 3 niveaux

Corollaire 3.1. [3][2][151] Soit \mathcal{WB}_n un graphe Cobweb d'ordre n , de taille m et de L niveau, comme nous l'avons défini avant. Donc, l'indice de Wiener de ce graphe est :

$$W(\mathcal{WB}_n) = \begin{cases} \frac{3p^3L^2+4p^2L^3-4p^2L-3pL^2}{24}, & \text{si } p \text{ est impair} \\ \frac{3p^3L^2+4p^2L^3-4p^2L}{24}, & \text{si } p \text{ est pair} \end{cases} \quad (3.39)$$

Démonstration. Puisque les deux cas (pair et impair) sont similaires, nous allons nous limiter à démontrer le cas où p est impair, et l'autre cas est de la même façon. Donc, on utilise la troisième formule de l'équation 3.1, on a :

$$W(\mathcal{WB}_n) = \frac{1}{2} \sum_{u \in V(\mathcal{WB}_n)} w(u, \mathcal{WB}_n); \quad (3.40)$$

On utilise le lemme 3.9,

$$\begin{aligned} W(\mathcal{WB}_n) &= P \frac{\sum_{l=1}^L (pl(l-1) + \frac{L(2Lp-4pl+(p+1)^2-2)}{4})}{24} \\ &= \frac{1}{24} (6L(2pL^2 + Lp^2 + 2Lp - L) + 4pL + 4pL(3L + 2L^2) - 12L(pL^2 + 2pL + p)) \\ &= \frac{1}{12} (3p^3L^2 - 3pL^2 + 4p^2L^3 - 4p^2L). \end{aligned}$$

□

Nous avons pu voir dans cette section l'importance des mathématiques dans toute la vie humaine. Et puisque l'informatique est le terrain d'application des mathématiques, nous allons essayer de donner à ce bagage mathématique un goût informatique dans nos prochaines recherches.

3 L'indice de Wiener pour les arbres

Les calculs directes utilisant la définition de l'indice de Wiener pourraient s'apparaître très faciles, surtout, lorsque le nombre de sommets du graphe considéré est petit. Mais ça va être presque impossible, dans le cas des grands graphes moléculaires, par ce que les calculs devient difficiles et impraticables. Tous ça exige des mathématiciens et informaticiens chimistes de doubler leurs efforts pour fournir des formules ou/et des méthodes qui vont facilitées les tâches aux chimistes.

Comme il est mentionné au chapitre I, le chemin entre deux sommets d'un arbre est unique, et par conséquence, le calcul de l'indice de Wiener d'un arbre serai relativement facile que d'autre graphe. Dans cette section nous allons donner un petit état de l'art des différentes formules, qui ont été utilisées pour le calculer. Pour réduire autant que possible ne nous allons pas inclure les démonstrations des formules verront dans cette section. En citant juste leurs références, Et pour ceux qui seraient en besoin peuvent les revoir aux références de formule ou en [168].

3.1 Les formules directes

La première formule que nous allons la voir dans ce chapitre, est une formule basique qui s'applique juste sur les arbres, et qui a été proposée aussi par H. Wiener [98] en 1947. Cette formule et contrairement à celle proposée lors de la première définition de l'indice de Wiener 3.1 comptant le nombre d'arêtes nécessaire pour aller chercher tous les sommets à partir de chaque sommets du graphe, elle compte le nombre de fois nécessaire pour passer sur toute arête du graphe. L'annonce de cette formule demande d'abord les définitions suivantes.

Définition 3.13. Soit $e = (u, v) \in E(T)$ une arête de l'arbre T . les sous-arbres T_u et T_v sont définis comme étant les composés connectés de l'arbre T contenant u et v , respectivement. Leurs ordres sont notés par $n_u(e) = |V(T_u)|$ et $n_v(e) = |V(T_v)|$ [168][14][109].

Théorème 3.15. [168][14][109] Soit T un arbre. Donc,

$$W(T) = \sum_{e=(u,v) \in E(T)} n_u(e)n_v(e) \quad (3.41)$$

Démonstration. Soit l'arête $e \in T$. Puisque T est un arbre et puisque les chemins reliant les sommets de l'ensemble T_u et T_v sont uniques, alors tout chemin d'eux doit contenir l'arête e . Par conséquent le $n_u(e)n_v(e)$ est exactement le nombre de fois où l'arête e existe dans le chemin de toute pairs de sommets de T . Et donc, la somme de toutes les arêtes de T n'est que l'indice de Wiener. Voir [168] [14][109], pour plus de détaille. \square

Définition 3.14. Soit $e = (u, v) \in E(G)$ une arête du graphe G . Les ensembles des sommets $B_u(e)$ et $B_v(e)$ du graphe G sont définis comme suit :

$$B_u(e) = \{x \in V(G) : d_G(x, u) < d_G(x, v)\} \quad (3.42)$$

$$B_v(e) = \{y \in V(G) : d_G(y, v) < d_G(y, u)\} \quad (3.43)$$

On désigne par $n_u(e) = |B_u(e)|$ et $n_v(e) = |B_v(e)|$, Leurs cardinaux respectivement [168].

Théorème 3.16. [168] Soit T un arbre d'ordre n . Donc,

$$W(T) = \frac{1}{4}[n^2(n-1) - \sum_{(u,v) \in E(T)} [d_T(v) - d_T(u)]^2] \quad (3.44)$$

Démonstration. Voir [168]. \square

Corollaire 3.2. [168] Soit T un arbre d'ordre n . Donc,

$$W(T) = \frac{1}{4}[n(n-1) + \sum_{v \in V(T)} deg_T(v)d_T(v)] \quad (3.45)$$

Démonstration. Voir [168]. \square

3.1.1 Les points de branchement et les segments

Les auteurs J. Doyle et J. Graver en [130] ont proposé, d'autres types de formules pour calculer l'indice de Wiener d'un arbre. Mais leurs formules nécessitent les définitions avec lesquels nous allons commencer ce paragraphe, pour les comprendre.

3.3 L'indice de Wiener pour les arbres

Définition 3.15. Soient G un graphe connexe, et trois sommets v_1, v_2 , et $v_3 \in V(G)$. On dit que v_1, v_2 , et v_3 sont collinears, s'ils sont ordonnés de tel sort que :

$$d_G(v_i, v_j) + d_G(v_j, v_k) = d_G(v_i, v_k) \quad (3.46)$$

Dans le cas inverse, on dit qu'ils sont non collinears, et on noté leur 3 – subsets par : $\tau(G)$ [130].

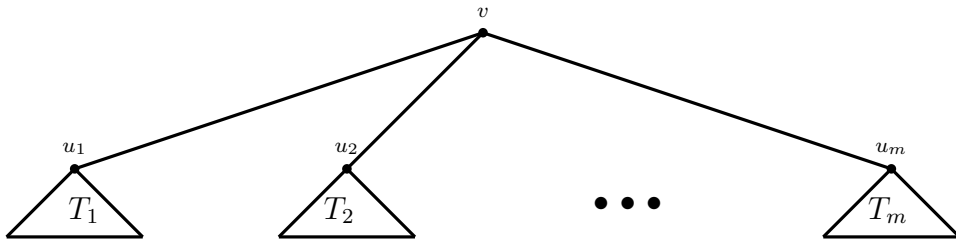


Figure 3.17 – Un arbre avec son point de branchement v et trois sous-arbre T_i

Le théorème désigné par, le théorème de Doyle-Graver est le suivant :

Théorème 3.17. [168][130] Soit T un arbre d'ordre n . Donc,

$$W(T) = C_3^{n+1} - \sum_{v \in V(T)} \sum_{1 \leq i < j < k \leq \deg_T(v)} |V(T_i)||V(T_j)||V(T_k)| \quad (3.47)$$

Où T_i, T_j et T_k sont les composés connexes de T .

Démonstration. Voir [168][130]. □

Dans la suite de ce petit paragraphe, nous allons annoncer d'autres résultats concernant le calcul d'indice de Wiener pour un arbre et qui utilise d'autres points de branchement et d'autres segments.

Théorème 3.18. [168][130] Soit l'arbre T d'ordre n . Donc,

$$W(T) = \sum_{S \text{ seg. de } T} n_1(S)n_{l_S+1}(S)l_S + \frac{1}{6} \sum_{S \text{ seg. de } T} l_S(l_S - 1)(3n - 2l_S + 1) \quad (3.48)$$

Où $n_1(S)$ et $n_{l_S+1}(S)$ sont les nombres de sommets des deux composés connexes obtenus par la suppression de tous les sommets internes de S et leurs arêtes correspondes.

Démonstration. Voir [168][13]. □

Théorème 3.19. [168] Soit T un arbre d'ordre n . Donc,

$$W(T) = \frac{1}{12}[(3n^2 + 1)(n - 1) - 3 \sum_{S \text{ seg. of } T} \frac{1}{l_S} [d_T(v_1) - d_T(v_{l_S+1})]^2 - \sum_{S \text{ seg. of } T} l_S^3] \quad (3.49)$$

Où v_1 et v_{l_S+1} sont les sommets terminaux de S .

Démonstration. Voir [168]. □

Définition 3.16. Soit T un arbre d'ordre n . On appelle l'étoile généralisé associé au sommet $v \in V(T)$, l'ensemble construit du sommet v et tout segment commençant de v . On note q_v la taille de cet étoile [168].

Remarque 3.1. [168] On note S^* le segment S moins 1, S^0 le segment S sans les sommets des extrémités, et on définit $SP(T)$ l'ensemble de tous les sommets terminaux des segments de l'arbre T .

Théorème 3.20. [168] Soit T un arbre d'ordre n . Donc,

$$W(T) = \frac{1}{12} \left[(3n + 1)(n - 1) + 3 \sum_{v \in SP(T)} q_v w(v, T) - \sum_{S \text{ seg. de } T} l_S^3 \right] \quad (3.50)$$

Où $w(v, T)$ est l'indice de Wiener du sommet v dans l'arbre T

Ce théorème que nous venons de l'annoncé, n'est que le théorème 3.19 récrit en fonction des segments et des étoiles généralisés.

Démonstration. Comme le sommet v est soit de degré 2 (Donc, il est un sommet intern du segment S) soit de degré plus grand que 2 (Donc, il est un sommet Terminal de S), donc la formule du Corollaire 3.2 peut être récrire comme suit :

$$\begin{aligned} W(T) &= \frac{1}{4} \left[n(n - 1) + \sum_{v \in SP(T)} \text{deg}_T(v) d_T(v) + \sum_{S \in T} \sum_{i=2}^{l_S} 2w(v_i, T) \right] \\ &= \frac{1}{4} \left[n(n - 1) + \sum_{v \in SP(T)} \text{deg}_T(v) d_T(v) + 2 \sum_{S \in T} \left(\frac{1}{2} (l_S - 1) [w(v_1, T) + w(v_{l_S+1}, T)] - \frac{1}{6} l_S (l_S^2 - 1) \right) \right] \\ &= \frac{1}{12} \left[3n(n - 1) + 3 \sum_{v \in SP(T)} \text{deg}_T(v) w(v, T) + 3 \sum_{S \in T} (l_S - 1) [w(v_1, T) + w(v_{l_S+1}, T)] - \sum_{S \in T} l_S^3 + \sum_{S \in T} l_S \right] \\ &= \frac{1}{12} \left[3n(n - 1) + 3 \sum_{v \in SP(T)} \text{deg}_T(v) w(v, T) + 3 \sum_{v \in SP(T)} (q_v - \text{deg}_T(v)) w(v, T) - \sum_{S \in T} l_S^3 + (n - 1) \right] \end{aligned}$$

$$= \frac{1}{12}[(3n + 1)(n - 1) + 3 \sum_{v \in SP(T)} q_v w(v, T) - \sum_{S \in T} l_S^3].$$

□

3.1.2 Les valeurs propres de la matrice Laplacienne

Parmi Les méthodes, de calcul de l'indice de Wiener d'un arbre, qui ont connues une large diffusion, surtout chez les mathématiciens d'algèbre, c'est celle qui utilise les valeurs propres de la matrice Laplacienne du graphe associé. Cette méthode, elle a été proposée en plusieurs sources depuis 1990, et c'est une méthode, qui diffère carèment des autres méthodes, et qui va ouvrir les portes devant l'indice de Wiener, pour qui il tire les intentions des chercheurs des autres disciplines, comme l'algèbre par exemple. Par ce que, elle va rendre l'indice de Wiener comme un point commun entre la chimie et l'algèbre, et même, elle va donner des astuces très importants, qui vont aider les algébristes dans leurs calculs délicats.

Théorème 3.21. [168] Soient T un arbre d'ordre n et $\lambda_1 \geq \dots \geq \lambda_{n-1} > \lambda_n = 0$ sont les valeurs propres de la matrice Laplacienne corresponde. Donc,

$$W(T) = n \sum_{i=1}^{n-1} \frac{1}{\lambda_i} \tag{3.51}$$

Démonstration. Voir [168].

□

3.2 Les méthodes récursives

Dans tous ce que nous venons de voir, on n'a pu calculer l'indice de Wiener pour les arbres que à l'aide des formules explicites. En d'autre terme, pour calculer l'indice de Wiener d'un arbre dans les formules précédentes, on a considéré tout l'arbre, sans prendre on considération ses sous composés connexes. Mais en réalité, il y a des arbres, qui sont des combinaisons de plusieurs autres simples arbres. Par exemple, on trouve des arbres qui sont la concaténation de plusieurs arbres étoiles, Paths ou d'autres. Donc, calculer l'indice de Wiener de l'arbre T , revient automatiquement à calculer celui de ses composés, et avec une simple opération de la somme ou la multiplication ou d'autres opération mathématique, on déduira l'indice de Wiener de l'arbre tout entier. On appelle ces méthodes de calcul, Les méthodes récursives. Durant ce petit paragraphe, nous allons essayer d'étudier, quelques combinaisons de quelques arbres, en donnant leurs formules pour calculer leurs indices de Wiener. On peut dire que la première astuce de la récursivité, est de supprimer feuille par feuille et à chaque fois on calcule l'indice de Wiener de chaque arbre résultant.

Théorème 3.22. [168] Soient T un arbre d'ordre $n \geq 2$, $v \in V(T)$ l'une de ses feuilles, $(u, v) \in E(T)$ l'une de ses arêtes et $T' = T - v$ l'un de ses sous graphes après la suppression de v de l'arbre T . Donc,

$$W(T) = W(T') + w(u, T') + n - 1 \tag{3.52}$$

Démonstration. Soient x, y deux sommets de T . Si le sommet supprimé est différent de x et y , c'est-à-dire $v \neq x$ et $v \neq y$, donc la distance entre x et y ne changera pas même après la suppression de v . Et par conséquent la somme de toutes ces distances est égale à l'indice de Wiener de T' . Maintenant, supposant que l'un de ces sommets x ou y égale à v , c'est-à-dire que $v = x$ ou $v = y$, donc $d_T(x, y) = d_{T'}(u, y) + 1$ où u est un sommet de T qui suit v . Donc, la somme de $n - 1$ pairs de $\{x, y\}$ égale à $w(u, T') + n - 1$, ce qui complète la démonstration. \square

Théorème 3.23. [168] Soient G un graphe connexe et $v \in V(G)$ le sommet sortant de G que l'on peut considérer comme une feuille de G . D'un autre côté soient $(v, u) \in E(G)$ une arête de G et $G' = G - v$ le sous graphe de G après la suppression de v . Donc,

$$W(G) = W(G') + w(u, G') + |V(G')| \tag{3.53}$$

Ce théorème est un cas particulier du théorème 3.22, par ce que le graphe de ce dernier est un arbre avec plusieurs feuilles, par contre le graphe du théorème 3.23 possède juste une seule feuille.

Démonstration. La démonstration de ce théorème on peut la trouver en [168]. En [68], il est démontré que ce théorème reste valable, en supprimant n'importe quel sommet du graphe. \square

Théorème 3.24. [168] Soit l'arbre T d'ordre $n \geq 2$, défini comme dans la Figure 3.17. Donc,

$$W(T) = \sum_{i=1}^m [W(T_i) + (n - |V(T_i)|)d_{T_i}(u_i) - |V(T_i)|^2] + n(n - 1) \tag{3.54}$$

Démonstration. Voir [168]. \square

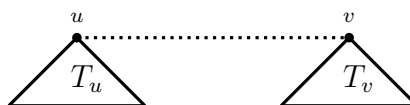


Figure 3.18 – Les arbres T_1 et T_2 , connexes par un k sommets

3.3 L'indice de Wiener pour les arbres

Théorème 3.25. [168][15] Soient T_u et T_v deux arbres avec $n_u = |V(T_u)|$, $n_v = |V(T_v)|$, $u \in V(T_u)$ et $v \in V(T_v)$ deux sommets. On construit l'arbre T utilisant les arbres T_u et T_v par la connection de u et v par un chemin de k nouveaux sommets (voir la Figure 3.18). Donc,

$$W(T) = W(T_u) + W(T_v) + (n_u + k)d_{T_v}(v) + (n_v + k)d_{T_u}(u) + (k + 1)n_u n_v + \frac{1}{2}(k^2 + k)(n_u + n_v) + \frac{1}{6}(k^3 - k) \quad (3.55)$$

Démonstration. Voir [168] et [15]. □

Corollaire 3.3. [168][15] Soient T_u et T_v deux arbres d'ordre n_u et n_v respectivement, et les sommets $u \in V(T_u)$ et $v \in V(T_v)$. Si T est construit de T_u et T_v par la connection du sommet $e = (u, v)$. Donc,

$$W(T) = W(T_u) + W(T_v) + n_u d_{T_v}(v) + n_v d_{T_u}(u) + n_u n_v \quad (3.56)$$

Démonstration. Ce résultat n'est que la formule du théorème 3.25 avec $k = 0$. □

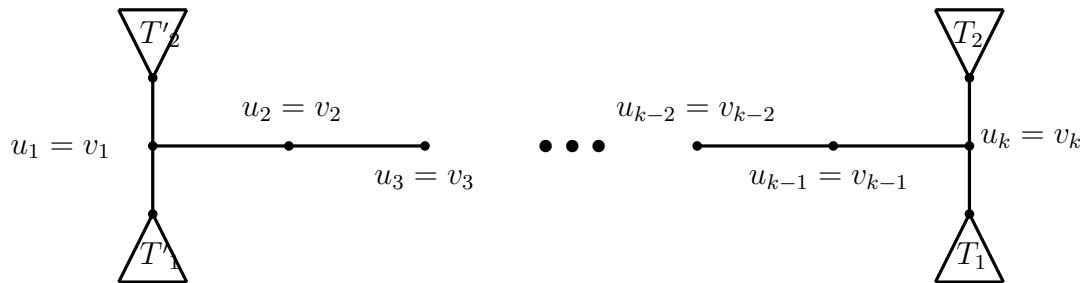


Figure 3.19 – Exemple d'identification des arbres

Théorème 3.26. [168][15] Soient T_1 et T_2 deux arbres avec $n_1 = |V(T_1)|$ et $n_2 = |V(T_2)|$. Et soient $p_1 = (u_1, u_2, \dots, u_k)$ un chemin en T_1 et $p_2 = (v_1, v_2, \dots, v_k)$ un chemin en T_2 , en élevant les points de branchement. Donc, l'indice de Wiener de l'arbre T , obtenu par l'identification de u_i et v_i , $i = 1, \dots, k$ (voir la Figure 3.19), peut être calculer, comme suit :

$$W(T) = W(T_1) + W(T_2) + (n_1 - k)d_{T_2}(v_1) + (n_2 - k)d_{T_1}(u_1) + 2(k - 1)[n_{u_k}(p_1) + n_{v_k}(p_2) - n_{u_k}(p_1)n_{v_k}(p_2)] - \frac{1}{2}k(k - 1)(n_1 + n_2) + \frac{1}{6}(k - 1)(5k^2 - k - 12) \quad (3.57)$$

Chapitre 3. L'indice de Wiener

Avec $n_{u_i}(p_i)$ est le nombre de sommets dans le composé connexe de T_i contenant $u_k(v_k)$ après la suppression de toutes les arêtes de p_i .

Démonstration. Voir [168] et [15]. □

Corollaire 3.4. [168] Soient T_1 et T_2 deux arbres d'ordre n_1 et n_2 respectivement et les sommets $u \in V(V(T_1))$ et $v \in V(V(T_2))$. Et soit T un arbre construit de T_1 et T_2 par l'identification de $e = (u, v)$. Donc,

$$W(T) = W(T_1) + W(T_2) + (n_1 - 1)d_{T_2}(v_2) + (n_2 - 1)d_{T_1}(v_1) \quad (3.58)$$

Démonstration. Ce corollaire n'est que la formule du théorème 3.26 pour $k = 1$. □

Lorsqu'on décortique Les précédentes formules, on constate qu'elles sont obtenues par la connexion de deux arbres, ce qui va nous pousser à penser aux arbres construits par la connexion de plus de deux arbres. Parmi les arbres dans cette situation on trouve, ce que l'on désigne par fasciagraphe F (voir la Figure 3.20). Le fasciagraphe F est l'arbre obtenu par la connexion de

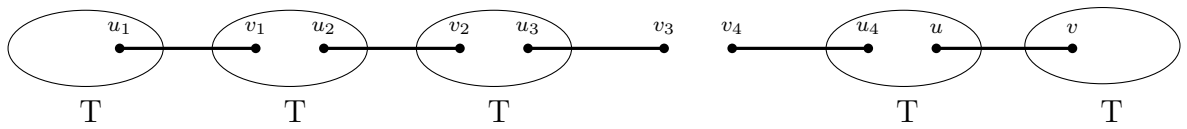


Figure 3.20 – Exemple du graphe fasciagraphe F

plusieurs copies de même graphe T dans une chaîne, de telle façon que le sommet u d'une copie de T est lié au sommet v de la copie T suivante par une arête (voir la Figure 3.20).

Théorème 3.27. [168][171][149] Soit F un fasciagraphe formé par m copies d'un arbre T d'ordre n , $m, n \geq 1$, et $u, v \in V(T)$ sont les sommets par lesquels les copies de T sont connectés. Donc,

$$W(F) = mW(T) + \frac{1}{2}nm(m-1)[d_T(u) + d_T(v)] + \frac{1}{6}n^2m(m-1)[(m-2)d_T(u, v) + m + 1] \quad (3.59)$$

Démonstration. Voir [168][171] et [149]. □

3.2.1 L'arbre de Thorn

Nous avons vu dans ce qui est passé la notion de la récursivité au sens d'augmentation de la valeur de l'indice de Wiener. Car lorsqu'on supprime, sommet par sommet dans un arbre

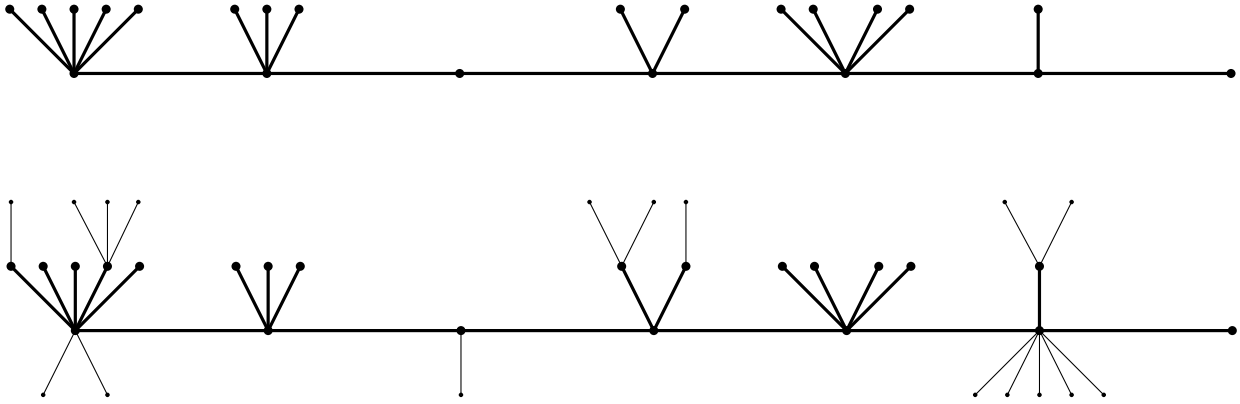


Figure 3.21 – Exemple d'un arbre et le graphe Thorn correspond

l'indice de Wiener s'augmente, ce qui est démontré dans les articles [148]. Exactement, mais dans le sens inverse, lorsqu'on ajoute, sommet par sommet à un arbre, la valeur de l'indice de Wiener se réduit. Donc pour passer de l'indice de Wiener d'un arbre T_1 à celui d'un autre arbre T_2 plus petit, il suffit d'ajouter quelques nouvelles feuilles au arbre T_2 [148].

Définition 3.17. [148] Soit T un arbre d'ordre n . Donc, T^* est appelé arbre de Thorn de T , si T^* se construit de T par l'ajoute de n_i nouveaux sommets, au sommet v_i de l'arbre T , avec $i = 1, 2, \dots, n$. (voir la Figure 3.21)

Remarque 3.2. [148] Il est important de noter que :

- Il est claire que le nombre de sommets de T^* est $n^* = n + \sum_{i=1}^n n_i$ et $deg_{T^*}(v_i) = deg_T(v_i) + n_i$
- Il faut noter aussi que, les deux arbre T et T^* , ne sont pas unique [168].

La formule suivante que nous allons la voir calculant l'indice de Wiener d'un arbre de Thorn T^* , en utilisant l'indice de Wiener de son arbre correspond T , était proposée par I. Gutman en 1998 [106].

Théorème 3.28. [168][106][117][108] Soit T un arbre d'ordre n et T^* son arbre de Thorn. Donc,

$$\begin{aligned}
 W(T^*) &= W(T) + \sum_{1 \leq i < j \leq n} (n_i + n_j) d_T(v_i, v_j) + \sum_{1 \leq i < j \leq n} n_i n_j d_T(v_i, v_j) + \left(\sum_{i=1}^n n_i \right)^2 + (n-1) \sum_{i=1}^n n_i \\
 &= \sum_{1 \leq i < j \leq n} (n_i + 1)(n_j + 1) d_T(v_i, v_j) + \left(\sum_{i=1}^n n_i \right)^2 + (n-1) \sum_{i=1}^n n_i
 \end{aligned}$$

Avec n_i est le nombre de nouveaux sommets connexes au sommet v_i .

Chapitre 3. L'indice de Wiener

Démonstration. Voir [168][106]. Nous avons aussi arrivé avec I. Gutman en [117][108], à une autre façon de le démontré, même si l'arbre considéré en [117][108], est un cas particulier du Thorn. Nous allons y revenir par la suite. \square

Corollaire 3.5. [168][117][108][106] Soient T un arbre des sommets v_1, v_2, \dots, v_n , et T^* sont arbre de Thorn avec $n_i = k$, $i = 1, 2, \dots, n$. Donc,

$$W(T^*) = (k + 1)^2 W(T) + nk(nk + n - 1) \quad (3.60)$$

Démonstration. On remplace dans la formule du théorème 3.28, on trouve :

$$\begin{aligned} W(T^*) &= W(T) + \sum_{1 \leq i < j \leq n} 2kd_T(v_i, v_j) + k^2 \sum_{1 \leq i < j \leq n} d_T(v_i, v_j) + \left(\sum_{i=1}^n k\right)^2 + (n-1) \sum_{i=1}^n k \\ &= (1 + 2k + k^2)W(T) + k^2 n^2 + kn(n-1) \\ &= (k + 1)^2 W(T) + nk(nk + n - 1). \end{aligned}$$

\square

Corollaire 3.6. [168][117][108][106] Soient T un arbre de n sommets et T^* sont arbre de Thron, avec les paramètres suivants $n_i = \deg_T(v_i)$, $i = 1, 2, \dots, n$. Donc,

$$W(T^*) = 9W(T) + (n-1)(3n-5) \quad (3.61)$$

Démonstration. Voir [168][117][108][106]. \square

Corollaire 3.7. [168][117][108][106] Soient T un arbre de n sommets, et T^* sont arbre de Thorn, avec les paramètres suivants, $n_i = m - \deg_T(v_i) \geq 0$, $i = 1, 2, \dots, n$. Donc,

$$W(T^*) = (m-1)^2 W(T) + [(m-1)n + 1]^2 \quad (3.62)$$

Démonstration. Voir [168][117][108][106]. \square

Définition 3.18. Lorsque l'arbre T est un Path, on appelle son arbre de Thorn T^* un caterpillar. Et, lorsque tous les sommets du caterpillar, qui ne sont pas des feuilles sont des sommets de degrés $a > 1$ ou bien $b > 0$, on le note par $T^*(a, b)$.

Corollaire 3.8. [168][106] Soit $T^*(a, b)$ un caterpillar d'un arbre T . Donc,

$$W(T^*(a, b)) = \frac{(a-1)^b}{6} [(a-1)(b-1)(b+7) + 6(a+1)] + 1 \quad (3.63)$$

Démonstration. On remplace dans la formule du corollaire 3.7 on a :

$$\begin{aligned}
 W(T'(a, b)) &= (a - 1)^2 W(P_b) + [(a - 1)b + 1]^2 \\
 &= (a - 1)^2 C_3^{b+1} + (a - 1)^2 b^2 + 2(a - 1)b + 1 \\
 &= (a - 1)b \left[(a - 1) \left(\frac{b^2 - 1}{6} + b \right) + 2 \right] + 1 \\
 &= \frac{(a - 1)b}{6} [(a - 1)(b^2 + 6b - 1) + 12] + 1 \\
 &= \frac{(a - 1)b}{6} [(a - 1)(b - 1)(b + 7) + 6(a + 1)] + 1
 \end{aligned}$$

□

3.2.2 L'arbre à k - sous divisions d'un arbre

En augmentant la longueur des segments dans un arbre, on obtient un autre concept d'arbre (voir [15]).

Définition 3.19. Soit T un arbre, l'arbre T' est appel k - sous divisions de T , si T' se construit de T , en remplaçant chaque arête de T par un Path de longueur $k + 1$. l'ordre de T' donc est $n' = k(n - 1) + n$ et le degré de chaque nouveau sommet est exactement égale à 2, et $deg_{T'}(v) = deg_T(v)$ pour tout $v \in V(T)$ [168][15].

Théorème 3.29. [168][15] Soient T un arbre d'ordre n et T' sont k - sous divisions. Donc,

$$W(T') = (k + 1)^3 W(T) + C_3^{n'+1} - (k + 1)^3 C_3^{n+1} \quad (3.64)$$

où n' est l'ordre de T'

Démonstration. On applique la formule du théorème 3.17 sur T' on a :

$$\begin{aligned}
 W(T') &= C_3^{n'+1} - \sum_{v \in V(T')} \sum_{1 \leq i < j < l \leq deg_{T'}(v)} |V(T'_i)| \cdot |V(T'_j)| \cdot |V(T'_l)| \\
 &= C_3^{n'+1} - \sum_{v \in V(T)} \sum_{1 \leq i < j < l \leq deg_T(v)} (k + 1)^3 |V(T_i)| \cdot |V(T_j)| \cdot |V(T_l)| \\
 &= C_3^{n'+1} + (k + 1)^3 W(T) - (k + 1)^3 C_3^{n+1}.
 \end{aligned}$$

□

Théorème 3.30. [168][15] Soit T un arbre de n sommets et $e = (u, v) \in E(T)$. et soit aussi T_1 un arbre, obtenu par un 1 – sous division de T . Donc,

$$W(T_1) = W(T) + \frac{1}{2}[d_T(u) + d_T(v) + n_u(e) + 2n_u(e)n_v(e) + n_v(e)] \quad (3.65)$$

Démonstration. Voir [168] et [15]. □

4 Le problème de l'inverse de Wiener

Comme nous l'avons déjà dit aux précédents chapitres les propriétés physico-chimiques des substances sont liées aux descripteurs topologiques en particulier l'indice de Wiener. Donc, il sera important de faire l'opération inverse, c'est-à-dire, de trouver les classes de graphes dont leurs indices de Wiener égaux à des nombres entiers. Proprement dit, **soit $w \in N$, quel(s) est (sont) le (s) graphe (s) G , que son(leurs) indice(s) de Wiener vérifie(nt) cette égalité : $W(G) = w$?** Cette problématique est connue sous le problème de l'inverse de Wiener ou «**La conjecture de Wiener**».

Les différentes réponses proposées depuis 1990 sont liées à des classes particulières de graphes. Parmi ces réponses, on trouve celle de I. Gutman, proposée en 1994 [122], qui concèrne les graphes connectés, et dans laquelle il a prouvé que tout nombre entier, outre que 2 et 5, peut correspond à un graphe connexe. Une année après, I. Gutman revient avec un autre résultat [121], mais cette fois ci, pour les graphes biparti. Sa dernière réponse exclue les nombres 2, 3, 5, 6, 7, 11, 12, 13, 15, 17, 19, 33, 37 et 39 des nombres entiers dont peuvent être des indices de Wiener pour des graphes biparti. Pour la classe des arbres, plusieurs propositions ont été faites [122][152][200][184][181]. Et pour plus de détails sur cette notion voir [168].

Ce chapitre nous a donné une idée sur l'indice de Wiener, cet indice qui se considère comme le père des indices topologiques. Dans ce chapitre nous avons discuté quelques outils mathématiques utilisés dans la littérature facilitant ses calculs, comme nous avons montré aussi les limites de ses outils. Ces limites vont s'apparaître clairement lorsqu'on compare ces outils avec notre nouvelle formule 3.5. On a montré dans ce chapitre aussi les avantages, les applications et même quelques perspectives, dont nous les reviendrons vers la fin de ce mémoire, de cette nouvelle formule. Et on a le terminé par une présentation rapide de **La conjecture de Wiener**. Dans le prochain chapitre, nous allons discuter d'autres indices (l'indice de degré de distance $DD(G)$, l'indice de Hyper-Wiener WW et l'indice de Terminal de Wiener $TW(G)$) qui ne sont que des extensions de l'indice de Wiener, et dont on a fait des contributions.

Chapitre 4

Les extensions de l'indice de Wiener

Au début de la fin des années soixante-dix, et comme nous l'avons déjà vu, l'indice de Wiener commence à affranchir les esprits des grands chercheurs de cette époque avant même leurs travaux. Il a connu une très bonne réputation dans ces années là, en raison de l'apparition d'une longue série des documents scientifiques qui le traite. Les années quatre-vingt-dix ont connu aussi, et après cette large diffusion de l'indice de Wiener chez la population chimiste, un phénomène qui a apparu un peu étrange à cette époque. Ce phénomène n'était que l'émergence d'un très grand nombre d'indices topologiques qui utilisent, soit les distances entre les sommets des graphes soit les degrés des sommets de ces graphes comme matière première de leurs définitions. Par conséquence, ces nouveaux invariants ont eu une très forte corrélation avec leur prédécesseur. Donc, durant cette époque et jusqu'à ces dernières années, la discipline de la théorie des graphes moléculaires a connu la création de plus de 10.000 nouvel indice topologique [163], et la majorité sera désignée après, par les extensions de l'indice de Wiener, comme par exemple, les indices de Hyper-Wiener, de degré de distance, et dernièrement l'indice de Terminal de Wiener, qui vont être le sujet de ce chapitre.

1 L'indice de Hyper-Wiener

Après le grand succès de l'indice de Wiener depuis les années 1970 et jusqu'à maintenant dans la discipline de chimie, de la pharmacologie et en chémo-informatique en générale, et surtout, lorsqu'on a découvert son importance dans le processus de la découverts des nouveaux médicaments, cette grandeur a connu plusieurs généralisations et même modifications, qui ont donné lieu à la création des nouveaux indices topologiques corespondents à des nouvelles propriétés moléculaires. La première généralisation de l'indice de Wiener a eu lieu en 1996, lorsque I. Gutman et ses co-auteurs en [104][114][107] ont récrit sa formule de définition sous forme de $d_G(k)$ (le nombre de paires de sommets ayant la même distance k). La définition de $W(G)$

devient donc comme suit :

$$W(G) = \sum_{\{u,v\} \subseteq V(G)} d(u,v) \quad (4.1)$$

$$= \sum_{k \geq 1} kd_G(k) \quad (4.2)$$

Ces équations dont la première est utilisée juste pour les graphes connexes, par contre la deuxième peut être utilisée pour n'importe quel graphe, même s'il est déconnexe [107]. Soit G le graphe formé des composés G_1, G_2, \dots, G_p , donc son indice de Wiener est :

$$W(G) = W(G_1) + W(G_2) + \dots + W(G_p)$$

1.1 L'invariant de Wiener-Type

L'indice de Wiener-Type ou l'invariant de Wiener-Type, comme il s'est désigné en [104][114][107], est considéré comme étant la fonction générale à partir de laquelle on fait sortir la définition de tout autre indice extension (successeur) de Wiener. Ces successeurs qui ne sont que les indices topologiques utilisant la distance entre les sommets du graphe en leurs définitions, comme les indices de Hyper-Wiener [104][114], de Harary [166] ... et même donc si leurs apparitions étaient en des années différentes et par des manières indépendantes leurs formules de définition sont semblables à l'invariant de Wiener-Type.

1.1.1 Définition

Soit λ un nombre réel ou complexe, l'invariant de Wiener-Typer pour les graphes connexes est défini en [107][114] comme étant la somme des multiplications de $d_G(k)k^\lambda$ du graphe G , c'est-à-dire :

$$W_\lambda(G) = \sum_{k \geq 1} d_G(k)k^\lambda \quad (4.3)$$

Et pour les graphes déconnexes [107] comme étant la somme des puissance de différentes distances du graphe G , c'est-à-dire :

$$W_\lambda(G) = \sum_{\{u,v\} \subseteq V(G) \times V(G)} (d_G(u,v))^\lambda \quad (4.4)$$

Malgré l'apparition de la formule de l'invariant de Wiener-Typer la première fois en 1996, ses cas particuliers sont considérés aux applications de la théorie des graphes chimiques plus avant et d'une manière indépendante. Par exemple pour $\lambda = -2$ et $\lambda = -1$ on trouve les indices

de Harary, défini en [56], et de l'inverse de Wiener défini en [166] apparus presque aux années quatre-vingt. Et dans le cas de $\lambda = \frac{1}{2}$ cet indice est apparu en [102] avant 1996. Nous allons étudier dans le reste de ce mémoire quelques autres indices pouvant être générés à partir de la formule de l'invariant de Wiener-Type, comme par exemple l'indice de Hyper-Wiener et de Terminal de Wiener.

1.1.2 Quelques propriétés de l'invariant de Wiener-Type

Nous avons vu au précédent chapitre que l'indice de Wiener de n'importe quel graphe simple, planaire et connexe $W(\mathcal{G})$ est majoré par l'indice de Wiener du graphe Path $W(\mathcal{P})$ et minoré par celui de graphe Maximale planaire $W(\mathcal{E})$ (voir la formule 3.4). Par la suite nous allons voir une paraît formule, mais concerne les arbre.

Définition 4.1. On appelle un graphe Star d'ordre n et de taille $m = n - 1$ noté \mathcal{S}_n , tout graphe dont $n - 1$ sommets sont adjacents au sommet du centre v_0 . Tous les sommets de ce graphe sont de degré 1 sauf le sommet central qui a le degré $n - 1$ (voir la Figure 4.1).

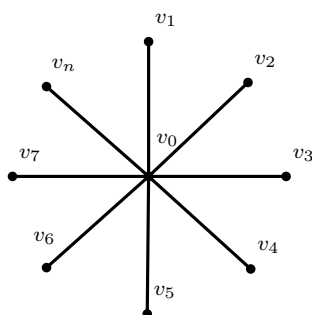


Figure 4.1 – Exemple de graphe Star \mathcal{S}_n d'ordre n

Le résultat suivant donne la valeur maximale et minimale de l'indice de Wiener d'un arbre \mathcal{S}_n d'ordre n .

$$W(\mathcal{S}_n) \leq W(\mathcal{T}_n) \leq W(\mathcal{P}_n) \tag{4.5}$$

Où \mathcal{S}_n est le graphe star d'ordre n , \mathcal{T}_n un arbre d'ordre n et \mathcal{P}_n est le graphe Path d'ordre n . Puisque l'invariant de Wiener-Type est la généralisation de l'indice de Wiener ainsi que d'autres indices, voici un résultat similaire à la formule 4.5 donné par I. Gutman en [104].

Théorème 4.1. [104] Soit T_n un arbre d'ordre n . Si $\lambda > 0$ donc :

$$W_\lambda(\mathcal{S}_n) \leq W_\lambda(\mathcal{T}_n) \leq W_\lambda(\mathcal{P}_n) \tag{4.6}$$

Théorème 4.2. [104][107] Soit T_n un arbre d'ordre n . Si $\lambda < 0$ donc :

$$W_\lambda(S_n) \geq W_\lambda(T_n) \geq W_\lambda(P_n) \quad (4.7)$$

Démonstration. Voir [104] et [107]. □

Prenant maintenant un arbre T d'ordre n et considérant l'arbre $T - e$ ($T - u$) l'arbre T en supprimant l'arête e (le sommet u) à condition de conserver sa connectivité. Donc, la relation entre l'invariant de Wiener-Type pour λ et $\lambda + 1$ est démontrée en [107] à travers les suivants théorèmes.

Théorème 4.3. [107] Soient T un arbre d'ordre n , $T - e$ l'arbre T en supprimant l'arête e et λ un nombre réel ou complexe, donc :

$$W_{\lambda+1}(T) = (n - 1)W_\lambda(T) - \sum_{e \in E(T)} W_\lambda(T - e) \quad (4.8)$$

Démonstration. Voir [107]. □

Théorème 4.4. [107] Soient T un arbre d'ordre n , $T - u$ l'arbre T en supprimant le sommet u et λ un nombre réel ou complexe, donc :

$$W_{\lambda+1}(T) = (n - 1)W_\lambda(T) - \sum_{u \in V(T)} W_\lambda(T - u) \quad (4.9)$$

Démonstration. Voir [107]. □

Considérant maintenant le composé connexe F d'une forêt (arbre non connecté composé de plusieurs composés connectés). L'indice de Wiener-Type de F est exprimé dans le théorème suivant.

Théorème 4.5. [107] Soient F un composé d'une forêt d'ordre n avec $m \leq n - 1$, et λ un nombre réel ou complexe, donc :

$$W_{\lambda+1}(F) = mW_\lambda(F) - \sum_{e \in E(F)} W_\lambda(F - e) \quad (4.10)$$

Démonstration. Voir [107]. □

1.2 L'indice de Hyper-Wiener

L'une des extensions de l'indice de Wiener, qui a connu depuis son émergence en 1993 par M. Randić une importance particulière pour les mathématiciennes chimistes, est l'indice de Hyper-Wiener. Originellement cet indice noté $WW(G)$ a été défini juste pour les graphes moléculaires non cycliques [156] avant qu'il sera généralisée sur tous les autres graphes. Et il a été considéré comme une généralisation de la formule de l'indice de Wiener pour les arbres que nous avons manipulé dans le chapitre III :

$$W(T) = \sum_{e \in E(T)} n_1(e|T)n_2(e|T) \quad (4.11)$$

Donc, selon Randić, si T est un arbre, et x, y deux sommets de T , on note $n_1(P_{xy}|T)$ et $n_2(P_{xy}|T)$ le nombre de sommets de T se trouvant dans les deux côtés du chemin P_{xy} [156]. L'indice de Hyper-Wiener de T a été défini comme suit :

$$WW(T) = \sum_{x < y} n_1(P_{xy}|T)n_2(P_{xy}|T) \quad (4.12)$$

Ici la sommation se fait sur toutes les paires de sommets de T , c'est-à-dire, sur tous les chemins de T .

Remarque 4.1. On remarque que, si la sommation de l'équation 4.12 restreint aux arêtes, on tombe dans la formule 4.11.

Mais avant cette définition et en 1990 Tratch, Stankevich et Zefirov ont proposé un nouvel indice topologique qui était similaire à l'indice de Hyper-Wiener [188], noté par WW_1 et nommé l'indice de Tratch-Stankevich-Zefirov, que nous l'on défini ainsi [156] :

$$WW_1(T) = \sum_{x < y} n_1(P_{xy}|T)d(x, y)n_2(P_{xy}|T) \quad (4.13)$$

Mais après, on a découvert que WW_1 n'est que le cas où $k = 0$ de la série suivante [156] :

$$WW_k(T) = \sum_{x < y} n_1(P_{xy}|T)(d(x, y))^k n_2(P_{xy}|T) \quad (4.14)$$

pour $k = 0, 1, 2, \dots$

1.2.1 Définition

Les indices de Hyper-Wiener et de Tratch-Stankevich-Zefirov vont se redéfinir par I. Gutman et ses co-auteurs en 1995 dans les articles [52] et [51], pour qu'ils s'adaptent à la définition de l'indice de Wiener-Type $W_\lambda(G)$ et pour qu'ils soient aussi utilisés pour n'importe quel graphe. Ces nouvelles définitions de $WW(T)$ et $WW_1(T)$ restent les seules utilisées jusqu'à maintenant.

Définition 4.2. [52] Soit G un graphe d'ordre n , on définit l'indice de Hyper-Wiener par :

$$WW(G) = \frac{1}{2}(W_1 + W_2) \quad (4.15)$$

Où W_1 et W_2 sont l'invariant de Wiener-Type avec $\lambda = 1$ et $\lambda = 2$ [52][51].

Définition 4.3. [51] Soit G un graphe d'ordre n , on définit l'indice de Tratch-Stankevich-Zefirov par :

$$WW_1(G) = \frac{1}{6}W_3 + \frac{1}{2}W_2 + \frac{1}{3}W_1 \quad (4.16)$$

Où W_1 , W_2 et W_3 sont l'invariant de Wiener-Type avec $\lambda = 1$, $\lambda = 2$ et $\lambda = 3$ [52][51].

Par la suite nous nous serons intéressés que par l'indice de Hyper-Wiener, dont on a fait des contributions publiées en différents articles.

1.2.2 La formule de l'indice d'Hyper-Wiener améliorée

Puisque la première définition de l'indice de Hyper-Wiener était pour les arbres, et vu l'unicité de chemin d'un arbre, l'indice de Hyper-Wiener ainsi que ses autres familles, ont été beaucoup appliqués sur les arbres, ce qui a donné lieu à un énorme résultats de ces indices topologiques pour les arbres. Nous avons voulu dépasser l'utilisation de ses indices avec les arbres aux leurs utilisations avec tous les autres types de graphes à condition qu'ils soient simples, connexes et planaires. Ce qui nous a poussé à proposer quelques nouvelles formules en fonction de $d_G(k)$ utilisées pour tout graphe simple, planaire, connexe et de n'importe quel diamètre, comme celle de l'indice de Wiener $W(G)$ vu précédemment, et deux autres seront étudiées dans ce chapitre l'une pour l'indice d'Hyper-Wiener $WW(G)$ et l'autre pour l'indice de degré de distance $DD(G)$. Ce paragraphe donc sera consacré premièrement pour la nouvelle réécriture de la formule de l'indice d'Hyper-Wiener $WW(G)$ puis à ses utilisations soit sur des graphes simples soit sur des graphes construits à partir des simples graphes.

1.2.3 La formule de l'indice d'Hyper-Wiener en fonction du diamètre

Nous commençons ce paragraphe par La formule de l'indice d'Hyper-Wiener $WW(G)$ que nous a récrit à partir de sa définition. Donc, cette formule n'est que la réécriture de la définition de l'indice d'Hyper-Wiener en fonction de $d_G(k)$, ce qui facilite le calcul de cet indice pour la plupart des graphes simples, connexes et planaires.

Théorème 4.6. [150][151][3][5][4][7] Soit G un graphe connexe, simple et planaire d'ordre n et de taille m . Donc l'indice d'Hyper-Wiener de ce graphe est :

$$WW(G) = \frac{1}{2}(3n(n-1) - 4m + (3^2 - 3)d_G(3) + \dots + (D^2 + D - 6)d_G(D)). \quad (4.17)$$

$$= \frac{1}{2}(3n(n-1) - 4m + \sum_{k=3}^D (k^2 + k - 6)d_G(k)). \quad (4.18)$$

Démonstration. Commençons par la définition de l'indice d'Hyper-Wiener :

$$WW(G) = \frac{1}{2}(W_1(G) + W_2(G))$$

En appliquant la formule de l'invariant de Wiener-Type pour $\lambda = 1$ et $\lambda = 2$.

On a $W_1(G) = W(G)$, Donc, il nous reste à calculer $W_2(G)$:

$$\begin{aligned} W_2(G) &= \sum_{k \geq 1} d_G(k)k^2 \\ &= d_G(1) + d_G(2)2^2 + \dots + d_G(D)D^2 \end{aligned}$$

En suivant la même démarche de la démonstration 2.1 on aura :

$$d_G(2) = \frac{n(n-1)}{2} - m - d_G(3) - \dots - d_G(D)$$

Et on remplace par $d_G(2)$ en $W_2(G)$, on aura :

$$\begin{aligned} W_2(G) &= m + d_G(2)2^2 + d_G(3)3^2 + \dots + d_G(D)D^2 \\ &= m + 2n(n-1) - 4m - 4d_G(3) - \dots - 4d_G(D) + d_G(3)3^2 + \dots + d_G(D)D^2 \\ &= 2n(n-1) - 3m + (3^2 - 4)d_G(3) + \dots + (D^2 - 4)d_G(D) \end{aligned}$$

□

Maintenant, et comme résultat directe pour les graphes de diamètre $D(G) = 2$, on aura la proposition suivante.

Proposition 4.1. [150][151][3][5][4][7] Soit G un graphe simple, connexe et planaire d'ordre n et de taille m et d'un diamètre $D(G) = 2$. Donc l'indice d'Hyper-Wiener est :

$$WW(G) = \frac{3}{2}n(n-1) - 2m. \quad (4.19)$$

Démonstration. La démonstration demande juste l'application de la formule 4.17 pour $D(G) = 2$ □

On utilisera dans la suite de ce paragraphe la formule 4.17 avec quelques catégories de graphes classés par diamètre.

1.2.4 L'utilisation de la formule avec des graphes à faible diamètre

Les graphes avec lesquels nous allons utiliser la formule 4.17 sont ceux que nous avons vu au précédent chapitre.

1. Les graphes de diamètre 2

L'utilisation de la formule 4.17 avec les graphes de diamètre 2 que nous avons défini au chapitre III (le graphe Fan noté \mathcal{F}_n , graphe Wheel noté \mathcal{W}_n , le graphe Maximale planaire noté \mathcal{E}_n , le graphe Butterfly noté \mathcal{B}_n et le graphe Crystal noté \mathcal{C}_n) donne leurs l'indice d'Hyper-Wiener à travers ce suivant théorème :

Théorème 4.7. [150][151] Soient les graphes Fan noté \mathcal{F}_n , Wheel noté \mathcal{W}_n , Maximale planaire noté \mathcal{E}_n , Butterfly noté \mathcal{B}_n et Crystal noté \mathcal{C}_n d'ordre tous n définis dans le précédent chapitre (voir les définitions 3.1, 3.2, 3.3, 3.4 et 3.5). Donc, les indices d'Hyper-Wiener de ces graphes sont :

G_n	$WW(G_n)$	n
\mathcal{F}_n	$\frac{3}{2}n^2 - \frac{11}{2}n + 6$	$n \geq 3$
\mathcal{W}_n	$\frac{3}{2}n^2 - \frac{7}{2}n$	$n \geq 3$
\mathcal{E}_n	$\frac{3}{2}n^2 - \frac{15}{2}n + 12$	$n \geq 4$
\mathcal{B}_n	$\frac{3}{2}n^2 - \frac{19}{2}n + 16$	$n \geq 6$
\mathcal{C}_n	$\frac{3}{2}n^2 - \frac{15}{2}n + 14$	$n \geq 5$

Démonstration. L'application du théorème 4.6 et les définitions de ces graphes suffit pour la démonstration de ce théorème. □

2. Les graphes de diamètre 3

On passe maintenant aux graphes de diamètre 3 qui seront les mêmes avec lesquels on a utilisé la réécriture de l'indice de Wiener du chapitre III. Considérons le graphe Double-Crystal noté \mathcal{DC}_n d'ordre n comme nous avons défini avant (voir la Figure 3.8).

Théorème 4.8. [5][4] Soit le graphe Double-Crystal noté \mathcal{DC}_n d'ordre n et de taille m . L'indice d'Hyper-Wiener de ce graphe est :

$$WW(\mathcal{DC}_n) = \frac{9}{4}n^2 - \frac{27}{2}n + 42, \quad \text{pour}(n \geq 10) \quad (4.20)$$

Démonstration. En appliquant juste le théorème 4.6, avec les informations du Lemme 3.1. □

3. Les graphes de diamètre 4

Nous allons réutiliser le théorème 4.6 avec les deux graphes de diamètre 4 sur lesquels on a travaillé dans le chapitre III (les graphes Sunflower noté \mathcal{FL}_n et Double-Wheel \mathcal{DW}_n d'ordre tous n définis précédemment (voir les définitions 3.8), 3.2).

Théorème 4.9. [5][150][151] Soit le graphe Sunflower noté \mathcal{FL}_n d'ordre n , défini avant. Donc, l'indice d'Hyper-Wiener de ce graphe est :

$$WW(\mathcal{FL}_n) = \frac{25}{8}n^2 - \frac{176}{8}n + \frac{151}{8} \quad \text{pour}(n \geq 11) \quad (4.21)$$

Démonstration. La démonstration de ce théorème demande l'application du théorème 4.6, avec les données du Lemme 3.2. □

On passe maintenant au graphe Double-Wheel \mathcal{DW}_n .

Théorème 4.10. [5][150][151] Soit le graphe Double-Wheel noté \mathcal{DW}_n d'ordre n et de taille m . Donc l'indice d'Hyper-Wiener de ce graphe est :

$$WW(\mathcal{DW}_n) = \frac{13}{4}n^2 - \frac{49}{2}n + 48, \quad \text{pour}(n \geq 10) \quad (4.22)$$

Démonstration. C'est juste l'utilisation du théorème 4.6 on prend en considérations les caractéristiques du graphe Double-Wheel du Lemme 3.3 ou bien de sa définition. □

4. Les graphes de diamètre supérieur à 4

Les graphes que nous voulons traiter comme graphes de diamètre supérieurs à 4 sont les graphes Path noté \mathcal{P}_n , Cycle noté \mathcal{O}_n et Grid noté \mathcal{G}_n (voir leurs définitions au chapitre III et les Figures 3.13, 3.14, 3.12).

– Le graphe Cycle :

Théorème 4.11. [3][150][151][7][6] Soit le graphe Cycle noté \mathcal{O}_n d'ordre $n \geq 2$ et de taille de m . Donc, l'indice d'Hyper-Wiener de ce graphe est :

$$WW(\mathcal{O}_n) = \begin{cases} \frac{n^2(n+2)(n+1)}{48}, & \text{si } n \text{ est pair} \\ \frac{n(n+3)(n^2-1)}{48}, & \text{si } n \text{ est impair} \end{cases} \quad (4.23)$$

Démonstration. Pour la démonstration on applique la formule du théorème 4.6, avec les informations contenues dans le lemme 3.4. \square

– Le graphe Grid

Théorème 4.12. [3][150][151][7][6] Soit le graphe Grid noté \mathcal{G}_n d'ordre n et de taille m . Donc, l'indice d'Hyper-Wiener de ce graphe est :

$$WW(\mathcal{G}_n) = \frac{n^4}{96} + \frac{n^3}{12} + \frac{5n^2}{24} + \frac{17}{3}n - 24 \quad \text{pour } (n \geq 4) \quad (4.24)$$

Démonstration. L'utilisation du théorème 4.6, avec les caractéristiques du graphe Grid contenues dans le lemme 3.5. \square

– Le graphe de Path

Théorème 4.13. [3][150][151][7][6] Soit le graphe Path noté \mathcal{P}_n d'ordre $n \geq 2$ et de taille m . Donc, l'indice d'Hyper-Wiener de ce graphe est :

$$WW(\mathcal{P}_n) = \frac{n^4 + 18n^3 - 41n^2 + 36n - 14}{24} \quad \text{pour } (n \geq 2) \quad (4.25)$$

Démonstration. Comme avant on applique le théorème 4.6 et le lemme 3.6. \square

2 L'indice de degré de distance

Comme l'indice d'Hyper-Wiener, l'indice de degré de distance a eu une grande variation en sa nomination et son expression, depuis 1989, avant qu'il devienne un indice standard avec un nom plus ou moins unifié. Cet indice est l'une des deux parties de la formule de l'invariant topologique désigné par *MTI* (Molecular topological index) apparu depuis 1989 par H. P. Schultz en [85], et qui a connu plusieurs modifications surtout par I. Gutman, qui le nommera après par l'invariant de Schultz.

L'indice de degré de distance connaît deux naissances différentes de deux parents différents dans deux pays différents. Et la surprise dans sa création c'est qu'elle a été faite dans le même journal et à la même date, mais dans de deux articles différents. En 28 Février 1994, Ivan Gut-

man soumet son article [103] intitulé «Selected properties of the Schultz Molecular Topological index» au journal Chem. Inf. Compu. Sci, dans lequel il développe la formule de *MTI* pour qu'il découvre dans celle-ci un nouvel indice topologique. En même date, A. A. Dobrynin et son co-auteur A. A. Kochetova, soumettent eux aussi leur article [8], au même journal, intitulé «Degree distance of a graphe : a degree analogue of the Wiener index» dans lequel, ils ont traité un nouvel indice avec le nom «degré de distance» avec quelques unes de ses propriétés. Le journal publiera les deux articles successifs en son volume 34.

Dans cette section, nous allons examiner l'indice de degré de distance, en détaillant, dans sa première partie, son prédécesseur, *MTI*, et ses propriétés, avant que nous passons à l'indice de degré de distance et ses modifications.

2.1 L'indice topologique moléculaire de Schultz (MTI)

En 1989, H. P. Schultz en [85], proposait un nouvel indice topologique nommé par l'indice topologique moléculaire et désigné par *MTI*. Cinq ans après sa création, l'indice topologique moléculaire devient le sujet d'étude de plusieurs chercheurs soit ses créateurs, H. P. Schultz, E. B. Schultz et T. P. Schultz en [86–91], soit autres auteurs, comme en [53, 57, 139, 199, 203].

2.1.1 Définition

L'indice topologique moléculaire, ou l'indice de Schultz comme l'appelle I. Gutman en [103] de temps en temps, est défini comme suit :

Définition 4.4. [86–91] Soient G un graphe moléculaire d'ordre n , A sa matrice d'adjacente, D sa matrice de distance, et $V = (deg(v_1), deg(v_2), \dots, deg(v_n))$ le vecteur de degrés de ses sommets. Donc on définit l'indice topologique moléculaire par :

$$MTI(G) = \sum_{i=1}^n [V(A + D)]_i \quad (4.26)$$

Remarque 4.2. [103] Pour simplifier la formule 4.26, on aura besoin des remarques suivantes :

1. On peut remplacer le vecteur de degrés des sommets du graphe G par :

$$V = (1, 1, \dots, 1)A \quad (4.27)$$

2. On rappelle que, si m est le nombre des arêtes de G :

$$\sum_{i=1}^n (deg(v_i)) = 2m$$

3. Désignant par P_2 le nombre de chemins de longueur 2 (voir l'équation de Wiener pour le calcul des point d'ébullition ?? du chapitre III), donc :

$$P_2 = \sum_{i=1}^n C_2^{deg(v_i)}$$

2.1.2 Quelques propriétés de MTI

En tenant compte de l'équation 4.27 et en développant la définition de MTI 4.26, I. Gutman trouve deux termes :

$$MTI(G) = \sum_{i=1}^n \sum_{j=1}^n [(A^2 + AD)]_{ij} = M + S \quad (4.28)$$

Où M n'est que le premier indice de Zagreb étudié depuis 1975 [111] et sur lequel nous avons déjà parlé dans nos précédents chapitres (voir le chapitre I et III par exemple) :

$$M = M_2 = \sum_{i=1}^n \sum_{j=1}^n [A^2]_{ij} = \sum_{i=1}^n (deg(v_i))^2$$

Mais, selon l'auteur de [103], le deuxième terme de l'équation 4.28 est une nouvelle notion, qui signalée en [88][203] avec la notation MTI' . Cette nouvelle notion selon Gutman, n'est que l'indice de Schultz (degré de distance) que nous détaillerons par la suite.

$$S = \sum_{i=1}^n \sum_{j=1}^n [AD]_{ij} = \sum_{i=1}^n \sum_{j=1}^n [DA]_{ij}$$

Parmi les propriétés dont MTI est connu, par sa relation avec l'indice de Wiener, est la propriété suivante :

Proposition 4.2. [53] Soient T un arbre d'ordre n et P_2 le nombre des chemins de longueur égale à 2 (voir la Remarque 4.2). Donc,

$$MTI(G) = 4W(G) + 2P_2 - (n - 1)(n - 2) \quad (4.29)$$

Démonstration. Voir l'article [53]. □

2.2 L'indice de degré de distance $DD(G)$

Comme on a pu le voir dans le précédent paragraphe, le développement de l'équation 4.28 a donné lieu à deux termes M et S . Nous avons identifié le premier terme, M , par le **premier indice de Zagreb**, et on a dit que le deuxième est l'expression d'un nouvel indice que quelques

chercheurs le nomme par, l'indice de Schultz, et qui le définit premièrement par 2.1.2, puis et après ses simplifications, il arrive à le définir comme suit :

$$S(G) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n [deg_{v_i} + deg_{v_j}] d(v_i + v_j) \quad (4.30)$$

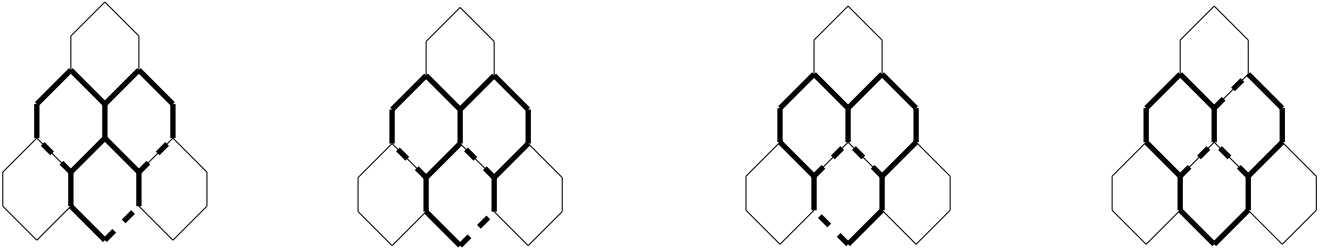


Figure 4.2 – L'arête créant les sous-graphes F_i de même nombre de sommets et différentes $DD(F_i)$

2.2.1 Définition

Au moment où I. Gutman arrive à faire sortir l'expression, qu'il a considérée comme un nouvel indice topologique, indice de Schultz, sans donner aucunes autres informations sur cet indice sauf quelques unes de ses propriétés purement mathématiques. Les auteurs de [8] ont donné le vrai nom de cette expression qui sera utilisée par la suite par tous les prochains chercheurs. Messieurs Dobrynin et Kochetova ont proposé l'indice de degré de distance, comme étant celui qui décrit l'arête créant des sous-graphes de mêmes nombre de sommets dans un graphe. autrement dit, considérant l'arête craint les sous-graphes marqués par des lignes en gras de la Figure 4.2. Selon Dobrynin et son collègue Kochetova, la localisation d'un sous-graphe $F_i \subseteq G$ peut être décrire par $DD(F_i)$ [8]. Donc, on peut dire que le vrai créateur de cet indice est Dobrynin et son collègue Kochetova.

Par analogie à la formule ?? de l'indice de Wiener pour les arbres, on trouve dans [103] aussi la définition de l'indice de degré de distance pour les arbres.

Définition 4.5. Soient T un arbre d'ordre n , et $n_u(e)$, $n_v(e)$ le nombre de sommets des deux côtes de l'arête e . Donc le l'indice de degré de distance de cet arbre est :

$$DD(T) = \sum_e [n_u(e) \sum_u deg(v_i) + n_v(e) \sum_v deg(v_i)] \quad (4.31)$$

Où, \sum_u et \sum_v sont les sommes sur tous les sommets de deux parties de l'arête e .

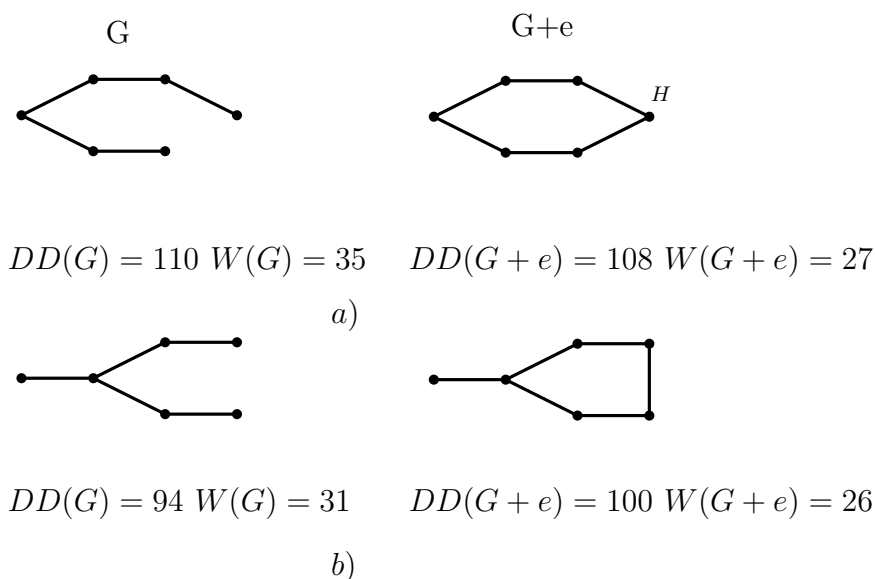


Figure 4.3 – Le graphe montrant la non-monotonie de $DD(G)$ vis-à-vis l'ajout d'une arête

Remarque 4.3. [8] Voici quelques remarques qui seront utiles par la suite :

1. Tous les sous-graphes F_i de la Figure 4.2 ont différentes valeurs de l'indice de degré de distance $DD(F_i)$
2. On peut récrire l'indice de degré de distance par :

$$\begin{aligned}
 DD(G) &= \frac{1}{2} \sum_{uv \in E(G)} (w(u, G) + w(v, G)) \\
 &= \frac{1}{2} \sum_{u \in V(G)} deg(u)w(u, G)
 \end{aligned}$$

3. On désigne par l'indice de degré de distance du sommet u la valeur suivante :

$$dd(u, G) = deg(u)w(u, G) \tag{4.32}$$

4. Les propriétés de l'indice de degré de distance peuvent être liés à ceux de l'indice de Wiener.
5. Si on ajoute une nouvelle arête à un graphe, la valeur du nouveau graphe est une fonction décroissante. Ce qui n'est pas correcte pour l'indice de degré de distance (voir la Figure 4.3).
6. On trouve dans l'article [103], un indice proche à l'indice de degré de distance, nommé

par la modification de l'indice de Schultz, et défini comme suit :

$$DD^*(G) = \sum_{i=1}^n \sum_{j=1}^n [deg_{v_i} deg_{v_j}] d(v_i, v_j) \quad (4.33)$$

2.2.2 Quelques propriétés de $DD(G)$

L'une des propriétés les plus connues de l'indice de degré de distance $DD(G)$ est celle qui le lie à son prédécesseur, l'indice de Wiener.

Théorème 4.14. [103] Soit T un arbre d'ordre n , donc :

$$DD(T) = 4W(T) - n(n-1) \quad (4.34)$$

Démonstration. Voir [103] par exemple. □

Théorème 4.15. [103] Soient T un arbre d'ordre n , n_v et n_u définis dans la définition 4.5 donc :

$$DD(T) = 4 \sum_e n_u(e)n_v(e) - \sum_e [n_u(e) + n_v(e)] \quad (4.35)$$

Démonstration. Voir [103] par exemple. □

On trouve en [185] les deux suivantes propositions de l'indice de degré de distance d'un graphe G $DD(G)$.

Proposition 4.3. [185] Soit G un graphe d'ordre n , de taille m et de diamètre $D(G)$, donc :

$$4(n-1)m - M_1(G) \leq DD(G) \leq -(D(G)-1)M_1(G) + 2(n-1)mD(G) \quad (4.36)$$

On aura égalité si et seulement si $D(G) \leq 2$.

Proposition 4.4. [185] Soit T un arbre d'ordre n , de taille m et de diamètre $D(T)$, donc :

$$DD(T) = (n-1)[2D(T)(m-n+1) - n] + 4W(T) \quad (4.37)$$

Et parmi les travaux récents traitant l'indice de degré de distance on trouve en [172], le théorème suivant.

Théorème 4.16. [172] Soit G un graphe d'ordre n et de diamètre $D(G)$ donc :

$$DD(T) \leq \frac{1}{4}nD(G)(n-D(G))^2 + O(n^{\frac{7}{2}}) \quad (4.38)$$

2.2.3 La réécriture de la formule de l'indice de degré de distance

Le présent paragraphe va traiter l'indice de degré de distance $DD(G)$ de quelques classes de graphes en utilisant la réécriture de cet indice que nous donnerons sa démonstration d'une autre manière.

1. La formule de $DD(G)$ en fonction de $w(u, G)$

La formule que nous allons démontrer ici, est une formule qu'on peut trouver en plusieurs références par exemple [103].

Théorème 4.17. [150][151][3][5][4][7][103] Soit G un graphe simple, fini, non orienté, connexe, planaire, d'ordre n et de taille m . Donc, l'indice de degré de distance de ce graphe est :

$$DD(G) = \sum_{u \in V(G)} w(u, G) \deg(u)$$

Démonstration.

$$\begin{aligned} DD(G) &= \sum_{\{u,v\} \subseteq V(G)} (\deg(u) + \deg(v))d(u, v) \\ &= \frac{1}{2} \sum_{u \in V(G)} \sum_{v \in V(G)} (\deg(u) + \deg(v))d(u, v) \\ &= \frac{1}{2} \sum_{u \in V(G)} \sum_{v \in V(G)} (d(u, v)\deg(u) + d(u, v)\deg(v)) \end{aligned}$$

On a :

$$\sum_{u \in V(G)} \sum_{v \in V(G)} d(u, v)\deg(u) = \sum_{u \in V(G)} \sum_{v \in V(G)} d(u, v)d(v)$$

Donc,

$$\begin{aligned} DD(G) &= \sum_{u \in V(G)} \sum_{v \in V(G)} d(u, v)\deg(u) \\ &= \sum_{u \in V(G)} w(u, G)\deg(u). \end{aligned}$$

□

Le cas particulier de ce théorème est pour les graphes de diamètre inférieure ou égale à 2. Ce cas demande la définition des quelques notion comme le premier indice de Zagreb, défini comme suit.

Définition 4.6. Soit G un graphe planaire, simple, connexe et fini. On définit le premier indice de Zagreb comme suit :

$$M_1(G) = \sum_{u \in V(G)} (\deg(u))^2 \quad (4.39)$$

Proposition 4.5. [150][151][3][5][4][7] Soit G un graphe non orienté, simple, connexe, planaire, d'ordre n , de taille m et de diamètre $D(G) \leq 2$. Donc, l'indice de degré de distance de ce graphe est :

$$DD(G) = 4(n - 1)m - M_1(G)$$

Où $M_1(G)$ est le premier indice de Zagreb défini avant.

Démonstration. On utilisant la formule 4.17, on a :

$$\begin{aligned} DD(G) &= \sum_{u \in V(G)} w(u, G) \deg(u). \\ w(u, G) &= \sum_{v \in V(G)} d(u, v) \end{aligned}$$

Puisque $D(G) = 2$ donc :

$$\begin{aligned} w(u, G) &= \sum_{\substack{v \in V(G) \\ d(u, v) = 1}} d(u, v) + \sum_{\substack{v \in V(G) \\ d(u, v) = 2}} d(u, v) \\ \sum_{\substack{v \in V(G) \\ d(u, v) = 1}} d(u, v) &= \deg(u) \\ n &= \sum_{\substack{v \in V(G) \\ d(u, v) = 1}} d(u, v) + n_2 + 1 \end{aligned}$$

Où n_2 est le nombre de sommets de v de $d(u, v) = 2$, donc :

$$\begin{aligned} w(u, G) &= \deg(u) + 2((n - 1) - \deg(u)) \\ DD(G) &= \sum_{u \in V(G)} (\deg(u) + 2((n - 1) - \deg(u))) \deg(u) \\ &= 2(n - 1) \sum_{u \in V(G)} \deg(u) - \sum_{u \in V(G)} \deg(u)^2 \\ &= 4(n - 1)m - M_1 \end{aligned}$$

□

2. *L'utilisation avec quelques graphes simples*

Le calcul de l'indice de degré de distance des graphes de diamètre faible, surtout le diamètre 2, nécessite l'utilisation soit le théorème 4.17 soit la proposition 4.5. Et l'utilisation de la proposition 4.5 demande quelques caractéristiques du graphe.

– *Les graphes de diamètre 2*

Le lemme suivant donne les valeurs du premier indice de Zagreb, des graphes de diamètre 2 avec lesquels nous allons utiliser la proposition 4.5 comme les graphes Fan noté \mathcal{F}_n , Wheel noté \mathcal{W}_n , maximale planaire noté \mathcal{E}_n , Butterfly noté \mathcal{B}_n et Crystal noté \mathcal{C}_n .

Lemme 4.1. [150][151] *Soient les graphes Fan noté \mathcal{F}_n , Wheel noté \mathcal{W}_n Maximale planaire noté \mathcal{E}_n , Butterfly noté \mathcal{B}_n et Crystal noté \mathcal{C}_n qui sont tous des graphes d'ordre n , définis avant (voir le chapitre III). Donc,*

G_n	m	$M_1(G_n)$	n
\mathcal{F}_n	$2n - 3$	$n^2 + 7n - 18$	$n \geq 3$
\mathcal{W}_n	$2n - 2$	$n^2 + 7n - 8$	$n \geq 3$
\mathcal{E}_n	$3n - 6$	$2n^2 + 12n - 44$	$n \geq 4$
\mathcal{B}_n	$2n - 4$	$\frac{4}{3}n^2 + \frac{4}{3}n - 10$	$n \geq 6$
\mathcal{C}_n	$3n - 7$	$2n^2 - 8n - 38$	$n \geq 5$

Démonstration. Sont juste des simples calculs à la main. □

Théorème 4.18. [150][151] *Soient les graphes Fan noté \mathcal{F}_n , Wheel noté \mathcal{W}_n Maximale planaire noté \mathcal{E}_n , Butterfly noté \mathcal{B}_n et Crystal noté \mathcal{C}_n qui sont tous des graphes d'ordre n , définis avant (voir le chapitre III). Donc leurs indices de degré de distance sont :*

G_n	$DD(G_n)$	n
\mathcal{F}_n	$7n^2 - 27n + 30$	$n \geq 3$
\mathcal{W}_n	$7n^2 - 23n + 16$	$n \geq 4$
\mathcal{E}_n	$10n^2 - 48n + 68$	$n \geq 4$
\mathcal{B}_n	$\frac{20}{3}n^2 - \frac{76}{3}n + 26$	$n \geq 6$
\mathcal{C}_n	$10n^2 - 48n + 66$	$n \geq 5$

Démonstration. En utilisant la proposition 4.5 et le lemme 4.1. Ou bien, en appliquant directement le théorème 4.17, □

– *Les graphes de diamètre 3*

On traite dans ce petit paragraphe le graphe Double-Crystal noté \mathcal{DC}_n que nous avons vu déjà avant (voir la Figure 3.8).

Théorème 4.19. [5][4] Soit le graphe Double-Crystal noté \mathcal{DC}_n d'ordre n et de taille m donc l'indice de degré de distance de ce graphe est :

$$DD(\mathcal{DC}_n) = 13n^2 - 88n + 172, \quad \text{pour}(n \geq 10) \quad (4.40)$$

Démonstration. En appliquant juste le théorème 4.17 et en utilisant le lemme 3.1. \square

– **Les graphes de diamètre 4**

Pour les graphes de diamètre 4 nous allons traiter les graphes Sunflower noté \mathcal{FL}_n et Double-Wheel noté \mathcal{DW}_n .

Théorème 4.20. [5][150][151] Soit le graphe Sunflower noté \mathcal{FL}_n d'ordre n . Donc l'indice de degré de distance de ce graphe est :

$$DD(\mathcal{FL}_n) = \frac{21}{2}n^2 - \frac{115}{2}n + 47 \quad \text{pour}(n \geq 11) \quad (4.41)$$

Démonstration. On applique le théorème 4.17 et le lemme 3.2. \square

On passe au graphe Double-Wheel \mathcal{DW}_n .

Théorème 4.21. [4] Soit le graphe Double-Wheel noté \mathcal{DW}_n d'ordre n et de taille m . Donc l'indice de degré de distance de ce graphe est :

$$DD(\mathcal{DW}_n) = 11n^2 - 58n + 74, \quad \text{pour}(n \geq 10) \quad (4.42)$$

Démonstration. L'utilisation du théorème 4.17, et le lemme 3.3 suffit pour la démontrer. \square

– **Les graphes de diamètre supérieur à 4**

Pour les graphes ayant un diamètre supérieur à 4, on trouve le graphe Path noté \mathcal{P}_n , Cycle noté \mathcal{O}_n et Grid noté \mathcal{G}_n (voir les Figures 3.13, 3.14 et 3.12).

– **Le graphe Cycle :**

Théorème 4.22. [3][150][151][7][6] Soit le graphe Cycle noté \mathcal{O}_n d'ordre $n \geq 2$ et de taille de m . Donc l'indice de degré de distance de ce graphe est :

$$DD(\mathcal{O}_n) = \begin{cases} \frac{1}{2}n^3, & \text{si } n \text{ est pair} \\ \frac{1}{2}n^3 - \frac{1}{2}n, & \text{si } n \text{ est impair} \end{cases} \quad (4.43)$$

Démonstration. On applique le théorème 4.17 et le lemme 3.4. \square

– **Le graphe Grid :**

Théorème 4.23. [3][2][4][151] Soit le graphe Grille noté \mathcal{G}_n d'ordre n et de taille m . Donc l'indice de degré de distance de ce graphe est :

$$DD(\mathcal{G}_n) = \frac{n^3}{2} + \frac{n^2}{2} - 2n \quad \text{pour } (n \geq 2) \quad (4.44)$$

Démonstration. On applique la formule du théorème 4.17 et le lemme 3.5. \square

– **Le graphe Path :**

Théorème 4.24. [3][7] Soit le graphe Path noté \mathcal{P}_n d'ordre $n \geq 2$ et de taille m . Donc, l'indice de degré de distance de ce graphe est :

$$DD(\mathcal{P}_n) = \frac{n(n-1)(2n-1)}{3} \quad \text{pour } (n \geq 2) \quad (4.45)$$

Démonstration. L'utilisation juste du théorème 4.17 et le lemme 3.6 suffit pour sa démonstration. \square

3. L'indice de degré de distance du graphe Cobweb

Théorème 4.25. [7][6] Soit le graphe Cobweb noté \mathcal{WB}_n d'ordre n , de taille m et de L niveau, défini avant. Donc l'indice de degré de distance de ce graphe est :

$$DD(\mathcal{WB}_n) = 8W(\mathcal{WB}_n) - pw(u_{1,1}, \mathcal{WB}_n) - pw(u_{1,L}, \mathcal{WB}_n) \quad (4.46)$$

Où C_i est le graphe Cycle du niveau l_i , et u est un sommet de ce niveau.

Démonstration.

$$\begin{aligned} DD(\mathcal{WB}_n) &= \sum_{u \in V(\mathcal{WB}_n)} w(u, \mathcal{WB}_n) \deg(u) \\ &= w(u_{i,1}, \mathcal{WB}_n) \deg(u_{i,1}) + \dots + w(u_{i,L}, \mathcal{WB}_n) \deg(u_{i,L}) \\ &= 3(w(u_{i,1}, \mathcal{WB}_n) + w(u_{i,L}, \mathcal{WB}_n)) + 4 \left(\sum_{l=2}^{L-1} w(u_{i,l}, \mathcal{WB}_n) \right) \\ &= 3(pw(u_{1,1}, \mathcal{WB}_n) + pw(u_{1,L}, \mathcal{WB}_n)) + 4 \left(p \sum_{l=2}^{L-1} w(u_{1,l}, \mathcal{WB}_n) \right) \\ &= 4p \left(\sum_{l=1}^L w(u_{1,l}, \mathcal{WB}_n) \right) - pw(u_{1,1}, \mathcal{WB}_n) - pw(u_{1,L}, \mathcal{WB}_n) \end{aligned}$$

\square

4. L'indice de degré de distance du graphe Star-Graph

Définition 4.7. On appelle un graphe Star-Graph d'ordre n noté $G_{m,n}$ tout graphe connectant N graphes simples, connexes, planaires et non orientés d'ordre m_i où $i = 1, \dots, N$ par un sommet s . Ce graphe est d'ordre $n = \sum_i^N m_i$ et de taille m (voir la Figure 4.4).

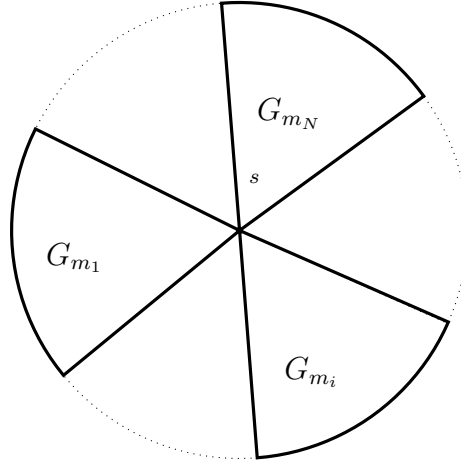


Figure 4.4 – Exemple du graphe Star-Graph $G_{m,n}$ d'ordre n

Théorème 4.26. [78] Soit le graphe Star-Graph noté $G_{m,n}$ d'ordre n et de taille m , défini avant. Donc, l'indice de degré de distance de ce graphe est :

$$DD(G_{m,n}) = \sum_{i=1}^N DD(G_{m_i}) - \sum_{i=1}^N w(s, G_{m_i}) deg_{G_{m_i}}(s) + \sum_{i=1}^N \left(\sum_{j=1 \text{ et } j \neq i}^N m_j - n + 1 \right) dd(s, G_{m_i}) +$$

$$w(s, G_{m,n}) deg_{G_N}(s) + \sum_{i=1}^N (w(s, G_{m,n}) - w(s, G_{m_i})) (2|E_{G_{m_i}}| - deg_{G_{m_i}}(s)) \quad (4.47)$$

Où $|E_{G_{m_i}}|$ est la taille du graphe G_{m_i} .

Démonstration. On utilise le théorème 4.17 on a :

$$\begin{aligned} DD(G_{m,n}) &= \sum_{u \in V(G_{m,n})} w(u, G_{m,n}) deg_{G_{m,n}}(u) \\ &= \sum_{u \in V(G_{m_1}) \cup \dots \cup V(G_{m_n})} w(u, G_{m,n}) deg_{G_{m,n}}(u) \\ &= \sum_{u \in V(G_{m_1} \setminus \{s\})} w(u, G_{m,n}) deg_{G_{m,n}}(u) + \dots + \sum_{u \in V(G_{m_N} \setminus \{s\})} w(u, G_{m,n}) deg_{G_{m,n}}(u) \\ &\quad + w(u, G_{m,n}) deg_{G_{m,n}}(s) + w(s, G_{m,n}) deg_{G_{m,n}}(s) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^N \sum_{u \in V(G_{m_i} \setminus \{s\})} \left[\left(\sum_{j=1 \text{ et } j \neq i} m_j - N + 1 \right) d(u, s) + w(s, G_{m,n}) - w(s, G_{m_i}) + \right. \\
 &\quad \left. w(u, G_{m_i}) \right] \text{deg}_{G_{m,n}}(u) + w(s, G_{m,n}) \text{deg}_{G_{m,n}}(s) \\
 &= \sum_{i=1}^N \left[\left(\sum_{j=1 \text{ et } j \neq i} m_j - n + 1 \right) dd(s, G_{m_i}) + (w(s, G_{m,n}) - w(s, G_{m_i})) (2|E_{G_{m_i}}| \right. \\
 &\quad \left. - \text{deg}_{G_{m_i}}(s)) + DD(G_{m_i}) - w(s, G_{m_i}) \text{deg}_{G_{m_i}}(s) \right] + w(s, G_{m,n}) \text{deg}_{G_{m,n}}(s)
 \end{aligned}$$

Et ça termine la démonstration. □

Pour plus de détaille et résultats sur ce type de graphes voir [1] et [77].

3 L'indice de Terminal de Wiener

Récemment, de nombreuses études traitent «La matrice de distance terminale» [31] [158] ou «la matrice de distance réduite» [62] d'un arbre. Soit l'arbre T d'ordre n possédant k feuilles (les sommets de degré égal à 1), notés v_1, v_2, \dots, v_k . La matrice de distance terminale notée TD de l'arbre T est une matrice carée de dimension k où ses entrées td_{ij} sont les distances $d(v_i, v_j)$ entre les feuilles de cet arbre.

Les matrices de distance terminale sont utilisées dans la modélisation mathématique des protéines ou des codes génétique [31] [158] [157]. Elles sont proposées aussi pour qu'elles soient les sources d'une classe des descripteurs des structures moléculaires [158][62]. Il est important de signaler que la matrice de distance terminale d'un arbre T détermine la matrice de distance entière de l'arbre T , et par conséquence détermine l'arbre tout entière[142].

Remarque 4.4. Soient les remarques suivantes :

- Pour tout arbre T , on a toujours $2 \leq k \leq n - 1$.
- Les seules arbres d'ordre n de $k = 2$ et $k = n - 1$ sont, respectivement, l'arbre Path (voir la Figure 3.13) et l'arbre Star (voir la figure 4.1).
- La matrice de distance terminale est définie juste pour les arbre.

3.1 Définition

Motivés par les dernières recherches sur les matrices de distance terminale et ses applications [31, 62, 142, 157, 158], les auteurs de [110] ont proposé en 2009 l'indice de Terminal de Wiener d'un arbre T noté par $TW(T)$. Il se définit comme suit :

Définition 4.8. Soient l'arbre T d'ordre n , qui possède $k \leq n$ feuilles et notons l'ensemble de ces feuilles par $V_1(T) \subseteq V(T)$. Donc, l'indice de Terminal de Wiener est défini comme suit :

$$TW(T) = \sum_{\{u_i, u_j\} \subseteq V_1(T)} d(u_i, u_j) \quad (4.48)$$

3.2 Quelques propriétés

Comme toujours, pour faciliter les calculs, on cherche des propriétés qui s'adaptent aux différentes situations. C'est le cas pour ce nouvel indice topologique. On cite dans cette section, les plus importantes propriétés de l'indice de Terminal de Wiener. En les classifie selon, si elles sont pour la maximisation ou la minimisation ou seulement des propriétés simples.

3.2.1 Les propriétés simples

Comme on a vu dans ce mémoire, on a deux formules pour calculer l'indice de Wiener, soit pour n'importe quel graphe, soit pour les arbres. Et comme l'indice de Terminal de Wiener est un successeur de ce dernier, lui aussi possède ces deux définitions. La première propriété proposée en [110] de $TW(T)$ est la suivante.

Théorème 4.27. [110] Soient T un arbre d'ordre n et de k feuilles et e son arête, on dénote par $p_1(e)$ et $p_2(e)$ le nombre des feuilles de l'arbre T se trouvent sur les extrémités de e . Donc, on peut récrire l'indice de Terminal de Wiener comme suit :

$$TW(T) = \sum_e p_1(e).p_2(e) \quad (4.49)$$

Où la sommation est sur toutes les arêtes de l'arbre T .

Démonstration. Voir [110] □

Remarque 4.5. Pour toute arête de l'arbre T , on a : $p_1(e) + p_2(e) = k$ et $p_1(e), p_2(e) \geq 1$. Ce qui va donner :

$$k - 1 \leq p_1(e).p_2(e) \leq \left(\frac{k}{2}\right)\left(\frac{k}{2}\right) \quad (4.50)$$

Mais, si l'arête e est une arête externe (attachée à une feuille), on aura : $p_1(e).p_2(e) = k - 1$.

Les auteurs en [20] ont donné l'indice de Terminal de Wiener du graphe Thorn que nous avons déjà vu dans le précédent chapitre (voir la Figure 3.21).

Théorème 4.28. [20] Soit G_p le graphe thorn obtenu par l'attachement de p_i terminaux sommets au sommet v_i d'un graphe connexe G , $i = 1, 2, 3, \dots, n$. Donc, si $p_i > 0$ pour $i = 1, 2, 3, \dots, n$:

$$TW(G_p) = 2 \sum_{i=1}^n C_2^{p_i} + \sum_{1 \leq i < j \leq n} p_i \cdot p_j (d(v_i, v_j) + 2) \quad (4.51)$$

Démonstration. Voir [20] □

Comme il est mentionné toujours en [20], la formule 4.51 reste valable, si on ajoute zero sommets, c'est-à-dire, même si $p_i = 0$. Mais si on ajoute à tous les sommets le même nombre de sommets on aura :

Corollaire 4.1. [20] Soit G_p le graphe Thorn défini en théorème 4.28 et si $p_1 = p_2 = \dots = p_n = p > 0$. Donc,

$$TW(G_p) = p^2 W(G) + pn(pn - 1) \quad (4.52)$$

Démonstration. Voir [20] □

Théorème 4.29. Soient G un graphe connexe d'ordre n et v_1, v_2, \dots, v_k ses feuilles. Considérons l'égalité suivante :

$$p_i = \begin{cases} p, & \text{si } i = 1, 2, \dots, k; \\ 0, & \text{si } i = k + 1, \dots, n. \end{cases}$$

pour $p > 0$. Donc,

$$TW(G_p) = p^2 TW(G) + pk(pk - 1) \quad (4.53)$$

Démonstration. Voir [20] □

3.2.2 Les limites de l'indice de Terminal de Wiener

Dans ce paragraphe nous allons discuter les limites de l'indice de Terminal de Wiener c'est-à-dire, sa valeur maximale et sa valeur minimale.

1. L'arbre d'indice de Terminal de Wiener minimal

Le graphe de Path P_n d'ordre n est un graphe possédant deux feuilles, et $n - 2$ sommets internes, et par conséquence on a pour tous les arêtes $p_1(e) = p_2(e) = 1$, donc : $TW(P_n) = (n - 1)(1 \times 1)$. Et puisque 1 est la valeur minimale possible pour le produit $p_1(e) \cdot p_2(e)$, donc $(n - 1)$ est la plus petite valeur possible de l'indice de Terminal de Wiener.

Théorème 4.30. [110] Pour tout arbre T d'ordre n on a $TW(T) \geq n - 1$. Et on aura $TW(T) = n - 1$ si et seulement si l'arbre T est un Path.

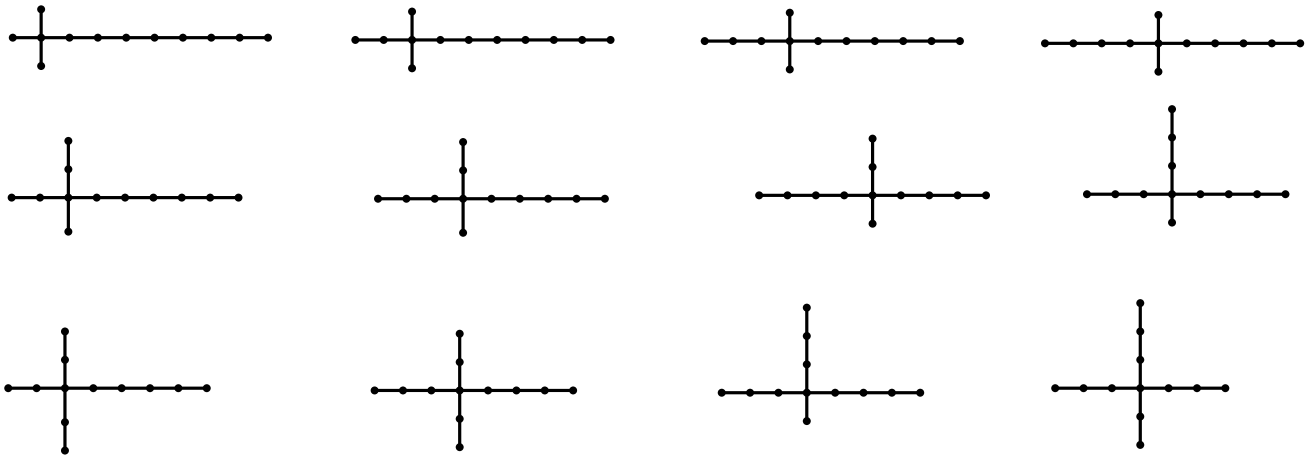


Figure 4.5 – Les arbres Starlike de degré 4 et d'ordre 12

2. Les arbres d'un nombre de nombre de feuilles fixe

Dans cette section nous traiterons les arbres d'un nombre de feuilles k fixe, qui possèdent donc k arêtes feuilles, ce qui exige la sommation de l'équation 4.49 k fois. Donc on peut récrire l'équation 4.49 comme :

$$TW(T) = k(k - 1) + \sum_{e'} p_1(e').p_2(e') \quad (4.54)$$

Où e' est l'arête interne de l'arbre T .

Avant de connaître les arbres possédant le l'indice de Terminal de Wiener minimal et maximal, on aura besoin de cette définition suivante.

Définition 4.9. On appelle arbre Starlike de degré k noté SL_k tout arbre possède un seul sommet de degré $k \geq 3$ et les autres sommets sont de degré inférieur ou égale à 2 (voir la Figure 4.5).

Théorème 4.31. [110] Les arbres Starlike d'ordre n et de degré $k \geq 3$ noté SL_k d'indice de Terminal de Wiener minimal $TW(SL_k) = (n - 1)(k - 1)$ sont parmi les arbres d'ordre n et d'un nombre feuilles fixe $k \geq 3$.

Démonstration. Pour plus de détail sur la démonstration de ce théorème voir [110]. \square

En appliquant ce theoreme, tous les douze arbres starlike d'ordre 12 de la Figure 4.5 auront la même valeur de l'indice de Terminal de Wiener égale à 33. Et comme conséquence de ce résultat. Voici une caractéristique de l'indice de Terminal de Wiener.

Corollaire 4.2. [110] *La puissance de la discrimination-Isomère de l'indice de Terminal de Wiener est très faible.*

Théorème 4.32. [110] *Tous les arbres d'ordre n où leurs arêtes internes e' vérifiantes : $p_1(e').p_2(e') = \binom{k}{2}\binom{k}{2}$ possèdent un indice de Terminal de Wiener maximal égale à : $TW = k(k-1) + (n-1-k)\binom{k}{2}\binom{k}{2}$. Ses arbres sont parmi les arbres d'ordre n possédant un nombre fixe de feuilles $k \geq 4$.*

Démonstration. Pour plus de détail sur la démonstration de ce théorème voir [110]. \square

3.2.3 L'arbre régulier T^* correspondant à un arbre T

Dans ce paragraphe, nous allons montrer notre contribution concernant l'indice de Terminal de Wiener [108]. Et pour la clarifier on aura besoin de notion suivante.

1. Définition

Définition 4.10. [108] Soient T un arbre d'ordre n noté T possède un sommet de degré maximal égale à Δ et un entier k , tel que $k \geq \Delta$. Notant u_1, u_2, \dots, v_n les sommets de T des degrés $deg_T(u_i)$, où $i = 1, 2, \dots, n$. On construit un arbre, noté T^* et nommé arbre régulier correspondant à l'arbre T , par l'attachement de $k - deg_T(u_i)$ feuilles à tout sommet u_i de T . Après la transformation tous les sommets de l'arbre T^* sont de degré k .

Exemple 4.1. Pour illustrer la définition 4.10 on considère l'exemple suivant : Soit T l'arbre de la Figure 4.6 de degré maximal $\Delta = deg_T(u_5) = 6$. Pour transformer l'arbre T à T^* , choisissant l'entier $k = 6$ et ajoutant à tout sommet v_i de T $6 - deg_T(u_i)$ feuilles. A la fin on aura l'arbre T^* de la même Figure 4.6 où tous ses sommets devient de degré égale à $deg(v_j) = k = 6$ (voir la Figure 4.6).

2. L'indice de Wiener de l'arbre régulier

Théorème 4.33. [108] *Soient T un arbre d'ordre n et de degré maximal Δ et T^* l'arbre régulier correspondant à T décrivent avant. Choississant k un entier tel que $k \geq \Delta$. Donc, l'indice de Wiener de T^* est :*

$$W(T^*) = (k-1)^2W(T) + 2kn\left(\frac{1}{2}kn - n + 1\right) + (n-1)^2 \quad (4.55)$$

Avant de donner la démonstration de ce théorème nous allons introduire certaines notations et faire la démonstration de quatre lemmes exploiteront dans la démonstration. Par la suite, on note T un arbre d'ordre n_T et T^* l'arbre régulier d'ordre $n_{T^*} = (k-1)n_T + 2$ correspondant à T . On notera aussi par $V(T_f^*)$ l'ensemble des feuilles de l'arbre T^* , $\Delta(\delta)$

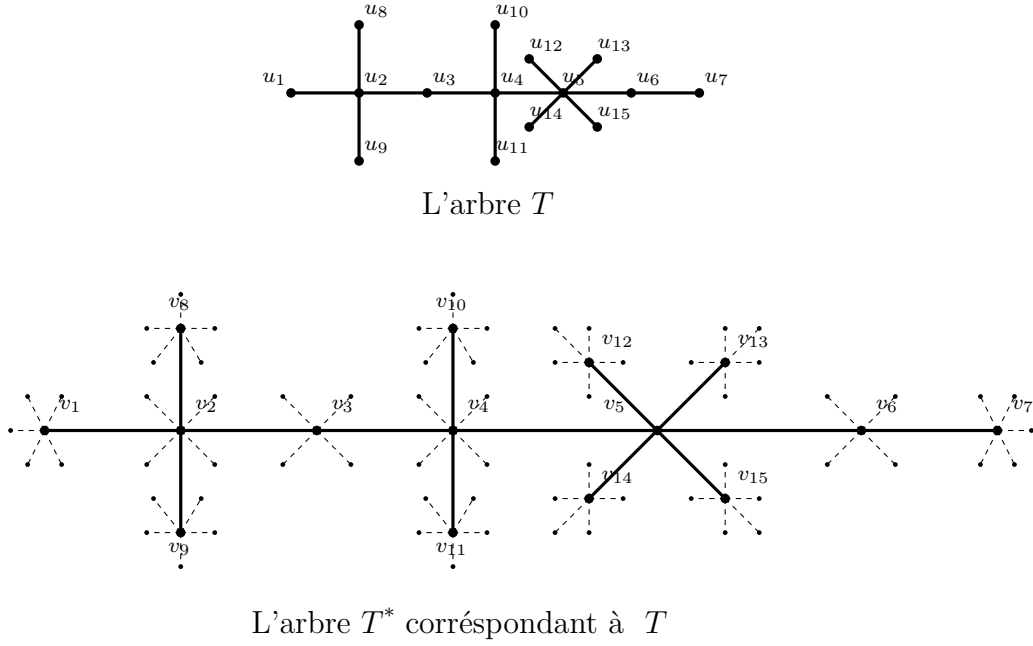


Figure 4.6 – L'arbre T et l'arbre régulier T^* correspondant à T pour $k = \Delta = 6$

le degré maximal (minimal) de l'arbre T , $deg_T(v)$ ($deg_{T^*}(v)$) le degré du sommet v de l'arbre T (l'arbre T^*) respectivement et on note par $w(v, T)$, $w(v, T^*)$ ou $w(v, T_f^*)$ l'indice de Wiener du sommet v dans l'arbre T , T^* , et T_f^* respectivement.

Lemme 4.2. [108] Soient T un arbre d'ordre n et T^* l'arbre régulier correspondant à l'arbre T . Pour tout $v \in V(T^*)$ on a :

$$w(v, T^*) = w(v, T) + w(v, T_f^*) \tag{4.56}$$

Démonstration. On sait que : $V(T^*) = V(T) \cup V(T_f^*)$, donc :

$$\begin{aligned} d_{T^*}(v) &= \sum_{u \in V(T^*)} d(v, u) \\ &= \sum_{u \in V(T)} d(v, u) + \sum_{u \in V(T_f^*)} d(v, u) \end{aligned}$$

□

Lemme 4.3. [108] Soient T un arbre d'ordre n et T^* l'arbre régulier correspondant à

l'arbre T . On a :

$$\begin{aligned} \sum_{u \in V(T_f^*)} \sum_{v \in V(T)} d(u, v) &= \sum_{v \in V(T)} \sum_{u \in V(T_f^*)} d(u, v) \\ &= kn^2 - n(n-1) + 2(k-2)W(T) \end{aligned}$$

Démonstration. La première égalité est évidente, pour la deuxième, on a :

$$\begin{aligned} \sum_{v \in V(T)} \sum_{u \in V(T_f^*)} d(u, v) &= \sum_{v \in V(T)} \sum_{u \in V(T_f^*)} (|V(T)| + d_T(v)) \\ &= \sum_{v \in V(T)} (k - \deg_T(v))(n + d_T(v)) \\ &= \sum_{v \in V(T)} (k.n + k.d_T(v) - n.\deg_T(v) - \deg_T(v).d_T(v)) \end{aligned}$$

Alors,

$$\begin{aligned} \sum_{v \in V(T)} (d_T(v)) &= 2W(T) \\ \sum_{v \in V(T)} (\deg_T(v).d_T(v)) &= 4W(T) - n(n-1) \end{aligned}$$

On remplace dans la précédente égalité on trouve le résultat. \square

Lemme 4.4. [108] Soient T un arbre d'ordre n et T^* l'arbre régulier correspondant à l'arbre T . On a :

$$W(T^*) = TW(T^*) + (2k-3)W(T) + (k-1)n^2 + n \quad (4.57)$$

Démonstration. Pour démontrer ce résultat, on utilise cette définition de Wiener :

$$\begin{aligned} W(T^*) &= \frac{1}{2} \sum_{v \in V(T^*)} w(v, T^*) \\ &= \frac{1}{2} \left(\sum_{v \in V(T_f^*)} w(v, T^*) + \sum_{v \in V(T)} w(v, T^*) \right) \\ &= \frac{1}{2} \left(\sum_{v \in V(T_f^*)} w(v, T_f^*) + \sum_{v \in V(T_f^*)} w(v, T) + \sum_{v \in V(T)} w(v, T) + \sum_{v \in V(T)} w(v, T_f^*) \right) \\ &= TW(T^*) + W(T) + \frac{1}{2} \left(\sum_{v \in V(T_f^*)} w(v, T) + \sum_{v \in V(T)} w(v, T_f^*) \right) \end{aligned}$$

En utilisant les Lemmes 4.2, 4.3, on a :

$$\begin{aligned} W(T^*) &= TW(T^*) + W(T) + kn^2 - n(n-1) + 2(k-2)W(T) \\ &= TW(T^*) + (2k-3)W(T) + (k-1)n^2 + n. \end{aligned}$$

Lemme 4.5. [108] Soit T et T^* deux arbres, définis avant. On a :

$$W(T^*) = \frac{1}{2}(TW(T^*) + (k^2 - 2)W(T) + k^2n^2 - kn^2 + 2kn - n + 1) \quad (4.58)$$

Démonstration. On utilise l'indice de degré de distance pour démontrer ce résultat :

$$\begin{aligned} DD(T^*) &= \sum_{v \in V(T^*)} \text{deg}_{T^*}(v)w(v, T^*) \\ &= \sum_{v \in V(T_f^*)} \text{deg}_{T^*}(v)w(v, T^*) + \sum_{v \in V(T)} \text{deg}_{T^*}(v)w(v, T^*) \\ &= \sum_{v \in V(T_f^*)} w(v, T^*) + k \sum_{v \in V(T)} w(v, T^*) \end{aligned}$$

En utilisant le Lemme 4.2, on aura :

$$\begin{aligned} DD(T^*) &= \sum_{v \in V(T_f^*)} (w(v, T) + w(v, T_f^*)) + k \sum_{v \in V(T)} (w(v, T) + w(v, T_f^*)) \\ &= 2TW(T^*) + 2kW(T) + \sum_{v \in V(T_f^*)} w(v, T) + k \sum_{v \in V(T)} w(v, T_f^*) \\ &= 2TW(T^*) + 2kW(T) + (1+k) \sum_{v \in V(T_f^*)} w(v, T) \end{aligned}$$

En utilisant, maintenant le Lemme 4.3, on aura :

$$\begin{aligned} DD(T^*) &= 2TW(T^*) + 2kW(T) + (1+k)(kn^2 - n(n-1) + 2(k-2)W(T)) \\ &= 2TW(T^*) + 2(k^2 - 2)W(T) + k^2n^2 + kn - n^2 + n \\ 4W(T^*) - n_{T^*}(n_{T^*} - 1) &= 2TW(T^*) + 2(k^2 - 2)W(T) + k^2n^2 + kn - n^2 + n \\ 4W(T^*) &= 2TW(T^*) + 2(k^2 - 2)W(T) + 2k^2n^2 - 2kn^2 + 4kn - 2n + 2. \end{aligned}$$

□

3. La démonstration du théorème 4.33

Démonstration. En utilisant les Lemmes 4.4 et 4.5, et on a :

$$\begin{cases} W(T^*) = TW(T^*) + (2k - 3)W(T) + (k - 1)n^2 + n \\ W(T^*) = \frac{1}{2}(TW(T^*) + (k^2 - 2)W(T) + k^2n^2 - kn^2 + 2kn - n + 1) \end{cases}$$

$$\begin{cases} \frac{1}{2}W(T^*) = \frac{1}{2}(TW(T^*) + (2k - 3)W(T) + (k - 1)n^2 + n) \\ W(T^*) = \frac{1}{2}(TW(T^*) + (k^2 - 2)W(T) + k^2n^2 - kn^2 + 2kn - n + 1) \end{cases}$$

$$\begin{aligned} \frac{1}{2}W(T^*) &= \left(\frac{k^2 - 2}{2} - \frac{2k - 3}{2}\right)W(T) + \frac{k^2n^2}{2} - \frac{kn^2}{2} + kn - n + \frac{1}{2} - \frac{(k - 1)n^2}{2} \\ &= \frac{1}{2}((k - 1)^2W(T) + 2kn(\frac{1}{2}kn - n + 1) + (n - 1)^2). \end{aligned}$$

□

4. L'indice de Terminal de Wiener de l'arbre régulier

Dans le théorème suivant on donne l'indice de Terminal de Wiener de l'arbre régulier $W(T^*)$ en fonction de son indice de Wiener $W(T)$.

Théorème 4.34. *Soient T un arbre d'ordre n et de degré maximal Δ et T^* l'arbre régulier correspondant à T décrivent avant. Choisisant k un entier tel que $k \geq \Delta$. Donc, l'indice de Terminal de Wiener de T^* est :*

$$TW(T^*) = (k - 2)^2W(T) + k^2n^2 - 3kn^2 + 2kn + 2n^2 - 3n + 1 \quad (4.59)$$

Démonstration. On utilise les Lemmes 4.4 et 4.5, et on a :

$$\begin{cases} W(T^*) = TW(T^*) + (2k - 3)W(T) + (k - 1)n^2 + n \\ W(T^*) = \frac{1}{2}(TW(T^*) + (k^2 - 2)W(T) + k^2n^2 - kn^2 + 2kn - n + 1) \end{cases}$$

$$\begin{aligned} \frac{1}{2}TW(T^*) &= \left(\frac{k^2 - 2 - 2(2k - 3)}{2}\right)W(T) + \frac{1}{2}(k^2n^2 - kn^2 + 2kn - n + 1 - 2(k - 1)n^2 - 2n) \\ &= \frac{1}{2}((k - 2)^2W(T) + k^2n^2 - 3kn^2 + 2kn + 2n^2 - 3n + 1). \end{aligned}$$

□

Comme nous avons vu durant tout ce chapitre, on a pu voir trois principaux indices topologiques ayant une relation forte avec son prédécesseur, l'indice de Wiener $W(G)$. On a vu les indices de

Degré de distance noté $DD(G)$, d'Hyper-Wiener noté $WW(G)$ et de Terminal de Wiener noté $TW(G)$. On a démontré les corrélations qui se trouvent entre eux, et on a vu leurs corrélations avec l'indice de Wiener $W(G)$. Dans ce présent chapitre aussi, on a vu les utilisations des indices de degré de distance et d'Hyper-Wiener avec certains graphes de différentes classes. Le chapitre suivant donnera une étude particulière des indices topologiques en générale concernant leurs corrélations entre eux d'une part et leurs utilisations comme moyens, surtout l'indice de Terminal de Wiener $TW(G)$, pour répondre à question fondamentale restée sans réponse plus d'un siècle. On discutera dans le chapitre V, la question suivante : pourquoi transformons-nous les molécules en des graphes ayant la représentation Kenograph et non à la représentation Plerograph ?

Troisième partie

La corrélation entre Plerograph et Kenograph

Chapitre 5

Les représentations Kenograph et de Plerograph

La structure d'une molécule organique peut être représentée par deux types des graphes moléculaires qui se désignent, selon A. Cayley, par **Plerogram** et **Kenogram**. Mais, dans la théorie des graphes chimiques contemporaine, presque toutes les molécules sont représentées par le type Kenogram. Ce qui a demandé aux chercheurs de la théorie des graphes chimiques de chercher le secret derrière cette domination, et d'essayer de revivre la terminologie de M. Cayley.

Dans le chapitre II, nous avons traité la notion de la modélisation et ses différentes étapes. L'étape la plus importante est entre autre celle de la sélection des variables (peu de descripteurs, variables dans le cas général, portent plus d'information sur l'activité ou sur la molécule de façon générale sera sans doute le but des différentes méthodes de la sélections des variables). En considérant ce principe et en ajoutant la molécule en discussion, nous avons cherché les différentes corrélations qui peuvent exister entre les différents indices topologiques en les regroupant selon la matière (distance, degré, diamètre, ...) de base de leurs définitions, au but d'éliminer ceux qui sont corélés. La relation entre les deux types (Kenogram et Plerogram) a été prise en considération afin d'utiliser celui le plus simple lors de calculer de l'un de ceux indices d'une part, et d'étudier leur utilisation parmi les facteurs d'influence aux méthodes de sélection de variables d'autre part.

Dans ce chapitre, nous allons étudier dans la première section les différentes évolutions des terminologies de la représentation des molécules depuis leur apparition, puis dans la deuxième section nous verrons la relation entre ces représentations. La troisième section traitera la corrélation entre les indices topologiques. A la fin de ce chapitre nous discuterons la relation existante entre les indices topologiques et les représentations moléculaires.

1 La terminologie des représentations graphiques

La représentation d'un problème par un dessin contribue souvent à sa compréhension. Le langage des graphes est construit, à l'origine, sur ce principe. Nombres de méthodes, de propriétés, de procédures ont été pensées ou trouvées à partir d'un schéma pour être ensuite formalisées et développées. Toutefois, entre cette notion vague où des points, représentant des individus, des objets, des lieux ou des situations, sont reliés par des traits, il y a une longue élaboration des concepts. La première difficulté à laquelle on peut être confronté concerne la terminologie. Dans notre vie quotidienne, la terminologie joue un rôle très important, par ce qu'elle décrit l'objet désigné par ce terme. Ce qui va le rendre unique dans les esprits de ses utilisateurs. Donc, l'unicité de terme, donne à l'objet en question une importance particulière, surtout chez les chercheurs, par ce qu'ils vont penser à la même chose et donc fournir le même résultat. Ce qui n'était pas le cas pour la transformation des graphes aux autres graphes désignés par quelques chercheurs par **les graphes de Thorns**. Comme nous allons le voir dans nos prochaines sous-sections, depuis 120 ans ou les chercheurs manipulent le même objets mais avec des termes différents ; et des fois ils fournissent le même résultat ce qui n'est pas bon au point de vue recherche scientifique, par ce que, c'est une perte de temps et des efforts des chercheurs.

1.1 Historique

Tout le monde s'accorde à considérer que la théorie des graphes est née en 1736 avec la communication d'Euler dans laquelle il proposait une solution au célèbre problème des ponts de Königsberg (Deux îles I_1 et I_2 sur la rivière Pregel à Königsberg, aujourd'hui désigné par Kaliningrad, étaient reliées entre elles ainsi qu'aux rivages R_1 et R_2 à l'aide de sept ponts P_1 , P_2 , P_3 , P_4 , P_5 , P_6 et P_7 , comme le montre la figure 5.1. Le problème posé consiste, à partir d'une terre quelconque I_1 , I_2 , R_1 et R_2 et à traverser chacun des ponts une fois et une seule et à revenir à son point de départ ; bien sûr sans traverser la rivière à la nage). Pendant les cent années qui suivirent, rien ne fut fait dans ce domaine de recherche.

En 1847, le physicien Kirchhoff développa la théorie des arbres pour l'appliquer à l'analyse de circuits électriques. Presque aux mêmes années, Cayley découvrit la notion d'arbre alors qu'il essayait d'énumérer les isomères saturés des hydrocarbures de type C_nH_{2n+2} . A cette époque, deux autres problèmes d'importance pour la théorie des graphes furent également proposés et partiellement résolus : **La conjecture des quatre couleurs** en 1840, et **le problème du chemin Hamiltonien** en 1859.

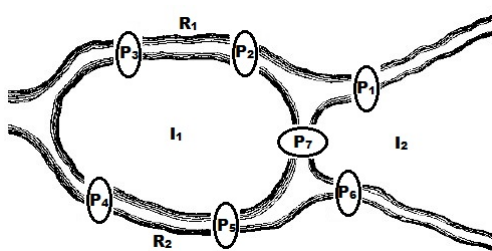


Figure 5.1 – La rivière Pregel et l'île de Kneiphof

Les années 1920 virent la résurgence de l'intérêt pour les graphes ; et à partir de 1946, la théorie des graphes a connu un développement intense sous l'impulsion de chercheurs motivés par la résolution de problèmes concrets.

Donc, la vraie naissance de la théorie des graphes chimiques est née en 1847, lorsque le mathématicien anglais distingué de la 19^{ème} siècle, Arthur Cayley (1821 – 1895), a proposé en son papier «On the mathematical theory of isomers» publié en 1874, deux types de graphe moléculaire qu'il nommait par Plerogram et Kenogram [10].

1.2 Les représentations Kenograph et Plerograph

1.2.1 Définitions

Comme nous l'avons déjà vu au premier chapitre, la représentation/type Plerogram, à la théorie des graphes chimiques contemporaine, est un graphe qui représente une molécule organique, et dans lequel tout atome est représenté par un sommet. Par contre une représentation/type dite Kenogram, est celle que l'on désigne par «graphe moléculaire à Hydrogène supprimé», en d'autres termes, c'est le graphe représentant une molécule organique dans lequel on ne représente par un sommet que les atomes du carbone (voir la Figure 1.5) [113][60][117][108].

1.2.2 Les vrais terms

Durant presque un siècle, et avec l'apparition de la théorie des graphes chimiques, toute la recherche a été faite sous ces deux types (Plerogram et Kenogram) ; en particulier le Kenogram. Et personne n'a pensé ni essayé à les renommer ni à les remplacer par d'autres graphes. La seule raison qui les a freinés et qu'on peut accepter c'est qu'à cette époque, comme nous venons de le voir, les chimistes n'ont pas encore découvert la notion des graphes ; et ils ont manipulé à sa place le mot diagramme. Le terme graphe n'a été manipulé que par les mathématiciens depuis 1736 ; et même la théorie des graphes à cette époque vécu (ces années) en silence. Mais

depuis 1920 et après l'édition de plusieurs livres traitant la théorie des graphes, les chimistes commencent à penser de changer les termes de Kenogram et Plerogram, car ils sont désignés aux graphes moléculaires, en composition de deux mots : **Keno** et **Plero** comme préfixes et **diagramme** comme suffixe [60].

En 1986 O.E.Polansky et I.Gutman en leur livre [120] intitulé «Mathematical Concepts in Organic Chemistry» proposaient la première modification de ces termes en **skeleton graphs** pour le Kenogram et **complete molecular graphs** pour le Plerogram. Ces nouveaux termes ne vont pas plier M. D.Vukičević et ses co-auteurs au papier [60] intitulé «Paths and Walks in acyclic structures : Plerographs versus Kenographs» qui vont profiter de l'occasion de la 75^{ème} anniversaire de M. Alexander T. Balaban pour le lui dédier en 2005. Cet article comporte la modification des nouveaux termes de Polansky et son collègue. Ils ont proposé **kenograph** pour Kenogram et **Plerograph** pour Plerogram. Cette modification est justifiée par le fait que l'objet résultant d'une molécule est un graphe et n'est pas un diagramme. Donc, pour eux A.Cayley n'était pas au courant du mot graphe, sinon il l'utiliserait. Et nous par la suite, nous garderons ces nouveaux termes, lorsqu'on veut désigner une représentation moléculaire.

1.2.3 La négligence du Plerograph

Les chercheurs utilisèrent ces termes tels quels, avec une large domination de la représentation Kenograph. Par ce que dans la vaste majorité des livres, papier, ... traitant la théorie des graphes chimiques, le terme **graphe moléculaire** pour eux veut dire Kenograph, et dans la littérature chimique contemporaine presque jamais le mot Plerograph est utilisé, ni un mot qui le signifie. A. Cayley dans son article [9], a justifié son choix du type Kenograph, même s'il est équivalent à celui de Plerograph dans le cas des énumérations des isomères des Alkanes, par ce qu'il contient moins de nombre de sommets que le type Plerograph et une structure simple et significative. Les chercheurs en [113][60][117][108] ont ajouté d'autres raisons de la domination du type Kenograph dans la littérature : comme l'apparition des indices topologiques qui sont presque tous définis sur le type Kenograph ; l'apparition de la théorie des graphes de Hückel [120][80] et sa connexion avec la théorie des graphes spectrales [120][80]. Et nous étudierons par la suite dans ce chapitre quelques très importantes autres raisons.

La représentation Plerograph n'était utilisée que dans quelques travaux de recherche scientifique limités, comme la modélisation des propriétés σ – *electron* des molécules organiques saturées [67, 75, 83, 160, 196].

Donc, si on veut résumer, on dit qu'en quelques applications chimiques le type Kenograph est obligatoire : la théorie des graphes de Hückel par exemple. En d'autres applications les

5.1 La terminologie des représentations graphiques

deux types sont équivalents. Par exemple : l'énumération des isomères des arbres en générale et les Alkanes en particulier. Mais le plus convivial au niveau pratique est la représentation Kenograph. Et il y a dans la théorie des graphes chimiques des domaines où il n'est pas évident de préférer le Kenograph au Plerograph. Par exemple lorsqu'on traite les invariants topologiques. Car ils s'utilisent aux études *QSAR/QSPR* qui se vérifieront après les tests sur les ordinateurs dans les laboratoire d'une manière pratique : c'est-à-dire on les expérimente sur des cellules biologiques ; ce qui signifie qu'il n'y a pas une raison claire de priorité d'indices topologiques du Kenograph sur ceux du Plerograph [113].

1.2.4 Le cas générale

Les représentations Kenograph et Plerograph ne sont que des cas particuliers pour une notion générale que nous avons déjà traité. La notion de **graphe Thorn** ou plus largement les **graphes Thorny** (voir la Figure 4.6). Le **graphe Thorn** est une notion proposée récemment par I.

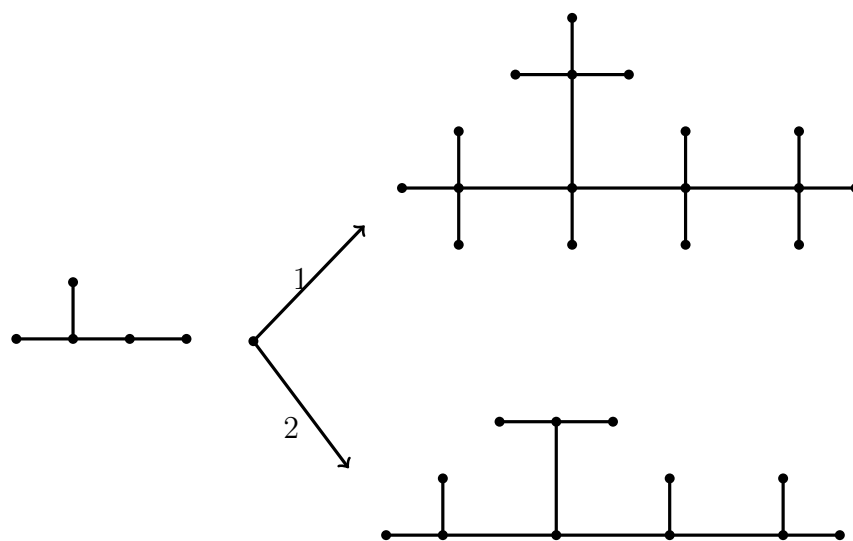


Figure 5.2 – Exemple de graphe Thorn pour $\alpha_{max} = 3$ pour 2 et $\alpha_{max} = 4$ pour 1

Gutman en [105][113], comme extension du concept généralisé du **graphe thorny** connu en avant. Ce concept introduit pour tout graphe G^* qui peut être obtenu à partir du graphe originale G par l'attachement de $p_i \geq 0$ nouvelles feuilles à chaque sommet i du graphe G qui doit être un arbre. Comme nous avons vu à la fin du précédent chapitre, des théorèmes ont été dérivés pour le cas spécial du **graphe Thorny** pour lequel $p_i = k - \delta_i$, où k est une constante qui doit être supérieure ou égale à Δ (le degré le plus élevé du graphe G), et δ_i est le degré du $i^{\text{ème}}$ sommet en G et évidemment G est un arbre (voir les Figures 5.2 et 4.6).

2 La relation entre les types Kenograph & Plerograph

Après la définition des deux types de représentations moléculaires, nous allons discuter l'utilisation exclusive du Kenograph *Ke* vis-à-vis le Plerograph *Pl*. Nous avons vu ensemble les raisons données par les chercheurs depuis les années 1990 traitant cette particularité; et nous nous sommes mis d'accord sur l'avantage du nombre réduit des sommets du Kenograph par rapport au Plerograph, ainsi que l'apparition des indices topologiques qui sont définis dès leur première apparition sur le Kenograph. On a aussi signalé le rôle que la théorie des graphes de Hückel a joué en ce sens. Dans cette section nous allons agrandir l'étude de cette problématique en ajoutant d'autres justificatifs qui peuvent être des avantages du Kenograph sur le Plerograph à travers les indices topologiques. Ces derniers, qui sont comme nous le savons les vrais descripteurs des structures moléculaires. Puisque les indices topologiques sont parmi les moyens les plus importants pour rendre les molécules manipulables dans la chémoinformatique en générale et le domaine de la découverte des nouveaux médicaments en particulier, donc si on arrive à trouver des corrélations entre les indices calculés sur le Kenograph et ceux calculés sur le Plerograph, cela veut dire que les deux structures sont équivalentes et on peut manipuler parmi eux, l'une la plus simple. Cette section aussi traitera le problème de la sélection des variables, pour les méthodes de modélisation; en sens de chercher les corrélations existantes entre plusieurs descripteurs (variables au jargon des méthodes de classifications) ce qui va aider les méthodes de sélection à réduire les descripteurs.

La section traitera donc trois parties, tournant toutes autour de la relation entre les types Kenograph et Plerograph, qu'on démontrera. Les deux dernières parties démontreront cette relation à travers la corrélation entre les indices topologiques calculés sur les deux types. L'avant dernière partie prend en compte les indices basés dans leur définition sur la distance et la dernière traitera les indices topologiques basés sur les degrés des sommets. La première partie donnera une autre façon de voir cette relation en utilisant les notions de Walks et les Paths dans le graphe.

2.1 La corrélation entre les Walks en Plerograph et Kenograph

Puisque les indices topologiques se basent sur les distances et les degrés, ce qui demande l'existence de la notion de **Paths** (le Paths est le sous-arbre dans lequel tous les sommets internes sont de degré égale à 2) et de la nouvelle notion de **Walk** introduite par les auteurs de [60] (qui est une séquence alternée de sommets et des arêtes, avec chaque arête étant incident sur les sommets précédant et succédant dans la séquence), nous allons présenter, donc, dans ce

5.2 La relation entre les types Kenograph & Plerograph

petit paragraphe introductif la relation entre le Plerograph Pl et Kenograph Ke à travers la corrélation entre leurs **Walks** des isomères d'Octane (c'est-à-dire C_8H_{18}), tirée de [60]. Notons

Octane	7-Walks en Kenograph	7-Walks en Plerograph
n-Octane	1254	29006
2-Methylheptane	1528	29834
3-Methylheptane	1676	30176
4-Methylheptane	1712	30212
3-Ethylhexane	1856	30554
2,2-Dimethylhexane	2284	31490
2,3-Dimethylhexane	2136	31328
2,4-Dimethylhexane	1994	31040
2,5-Dimethylhexane	1822	30680
3,3-Dimethylhexane	2602	32138
3,4-Dimethylhexane	2272	31652
3-Ethyl-2-methylpentane	2304	31688
3-Ethyl-3-methylpentane	2882	32750
2,2,3-Trimethylpentane	3072	33236
2,2,4-Trimethylpentane	2634	32372
2,3,3-Trimethylpentane	3218	33524
2,3,4-Trimethylpentane	2592	32462
2,2,3,3-Tetramethylbutane	4094	35072

Tableau 5.1 – La corrélation entre Les Walks du Ke et du Pl pour les octanes

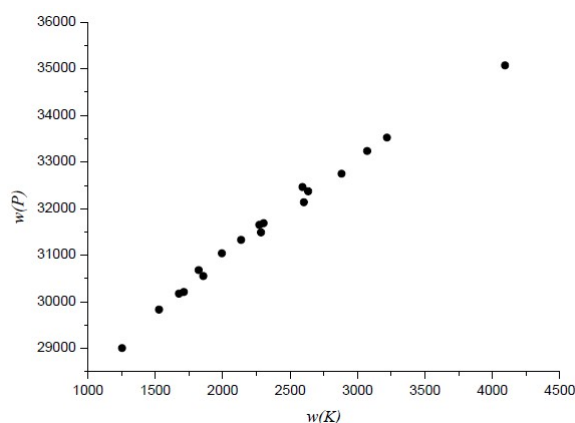


Figure 5.3 – La corrélation entre $w(Ke)$ et $w(Pl)$ pour les isomères d'Octane

par i – *Walks* le nombre de **Walks** de taille i . Le tableau 5.1 montre le nombre de 7 – *Walks*, c'est-à-dire le nombre de Walks de longueur 7.

Comme montre la Figure 5.3, la corrélation entre le walk de Plerograph $w(Pl)$ et celui de Kenograph $w(Ke)$ est presque linéaire, la chose confirmée par cette équation :

$$w(Pl) = 2.1w(Ke) + 26678.9 \quad (5.1)$$

Donc, cette corrélation qui a comme coefficient 0.995, a montré que si dans la modélisation *QSAR/QSPR* on utilise les Walks comme descripteurs des structures non-cycliques, il sera mieux d'utiliser le Kenograph à la place de Plerograph comme représentateur des molécules étudiées. Car même si la représentation Plerograph est la plus proche à la molécule au niveau forme, elle prendra plus du temps du calcul, car elle est plus complexe que celle de Kenograph.

2.2 La corrélation entre les *TI* basés sur la distance

Lorsqu'on a vu dans les précédents chapitres les relations entre les indices topologiques *TI* des graphes moléculaires et les propriétés physico-chimiques des molécules correspondantes, on s'est mis d'accord sur le fait que les indices topologiques représentent les molécules et qu'ils les caractérisent à l'aide des propriétés physico-chimiques. Ce qui rendra la classification des molécules possible. Et on a aussi vu l'augmentation exponentielle du nombre des indices topologiques. Et autant que ce nombre augmente autant qu'on aura mal à classer les molécules, ce qui exige, comme on a vu, l'existence des méthodes dites de sélection de variables (la deuxième étape du procédure de la classification). Ce paragraphe et le suivant vont nous donner quelques techniques qui joueront sûrement un rôle important dans ces méthodes. On va démontrer les différentes corrélations qui relient les indices topologiques entre eux en les regroupant selon les moyens de base de leur définition.

2.2.1 Le cas de l'invariant de Wiener-Type

1. *La formule générale*

On a dit précédemment que tous les indices topologiques basés en leur définition sur la distance ne sont qu'un cas particulier de l'invariant de Wiener-Type ($W_\lambda(G) = \sum_{k \geq 1} d_G(k)k^\lambda$) que nous avons déjà vu sa définition. En 1998 en article [113] il a été démontré qu'il existe une relation forte entre le Kenograph *Ke* et le Plerograph *Pl* à travers cet invariant.

Théorème 5.1. [113] *Soient les représentations Plerograph noté Pl, Kenograph noté Ke d'une molécule organique de type Alkane et λ un nombre réel ou complexe. Donc, l'invariant de Wiener-Type des deux représentations Ke et Pl peut s'exprimer comme*

5.2 La relation entre les types Kenograph & Plerograph

suit :

$$W_\lambda(Pl) = aW_\lambda(Ke) + b \quad (5.2)$$

Où a et b sont des nombres réels.

La formule 5.2 de ce théorème peut être, comme nous allons discuter à travers quelques figures et tableaux explicatifs de chaque indice, linéaire pour quelques indices topologiques (corresponds aux quelques différentes valeurs de λ) et approximative pour d'autres.

2. Les coefficients de corrélation de la formule générale

On commence la vérification du théorème 5.1 par la recherche des coefficients de corrélation qui peuvent lier la valeur de tout indice topologique basé sur la distance calculée sur le Plerograph et toute autre valeur du même indice calculée sur le Kenograph. Le

$\lambda \backslash n$		5	6	7	8	9	10
-4	a	0.44	0.50	0.53	0.56	0.59	0.62
	b	16.62	19.38	22.19	24.96	27.62	30.28
-3	a	0.82	0.93	1.00	1.05	1.12	1.18
	b	18.02	20.78	23.63	26.39	28.90	31.42
-2	a	1.54	1.72	1.83	1.93	2.05	2.14
	b	22.43	25.83	29.41	32.87	35.81	38.83
-1	a	2.82	3.09	3.25	3.38	3.55	3.67
	b	36.76	43.80	51.48	59.27	66.21	73.61
1	a	9	9	9	9	9	9
	b	256	361	484	625	784	961
2	a	15.53	14.55	14.01	13.64	13.26	12.99
	b	891.60	1385.14	1994.44	2729.26	3634.00	4680.33
3	a	26.10	22.74	20.96	19.81	18.71	17.94
	b	3366.55	5812.36	9040.43	13139.22	18608.24	25251.84
4	a	42.72	34.49	30.29	27.73	25.49	23.92
	b	13392.51	26002.15	44052.69	68332.49	103279.06	148241.84

Tableau 5.2 – Certaines valeurs de a et b de formule 5.2 en fonction de n et λ

Tableau 5.4, tiré de [113], nous donne les Coefficients de corrélation de l'équation 5.2, calculés pour les ensembles de tous les isomères C_nH_{2n+2} , pour les valeurs de $n = 5 - 10$. Ce tableau clarifie donc, la formule 5.2 et la confirme. Nous rappelons que tout indice topologique correspond a une valeur particulière de λ de la formule 5.2. Certains indices sont déjà liés à quelques valeurs, comme est déjà dit au précédent chapitre toute valeur $3 \geq \lambda \geq -2$ correspond à un indice topologique particulier (voir le tableau 5.3), par contre d'autres valeurs ne sont pas encore corresponds à aucun autre indice. Le Tableau

La valeur de λ	L'indice topologique correspond
-2	indice de Harary
-1	indice de l'inverse de Wiener
1	indice de Wiener
2 et 1	indice hyper de Wiener
3, 2 et 1	indice de Tratch-Stankevich-Zefirov

Tableau 5.3 – Les valeurs de λ et les indices correspond

$n \setminus \lambda$	-4	-3	-2	-1	1	2	3	4
5	0.9992	0.9992	0.9993	0.9996	1	0.9998	0.9989	0.9974
6	0.9979	0.9981	0.9986	0.9992	1	0.9998	0.9990	0.9979
7	0.9959	0.9965	0.9976	0.9987	1	0.9996	0.9985	0.9969
8	0.9942	0.9953	0.9968	0.9984	1	0.9996	0.9984	0.9965
9	0.9904	0.9928	0.9957	0.9982	1	0.9996	0.9986	0.9969
10	0.9885	0.9914	0.9950	0.9979	1	0.9996	0.9986	0.9969

Tableau 5.4 – les Coefficients de corrélation de formule 5.2 pour quelques λ

5.2, tiré aussi de [113], montre certaines valeurs des variables a et b de la même formule 5.2. Ce tableau explique aussi l'existence de ces réels dans la précédente formule exprimée par leurs dépendances des valeurs de λ et de n (le nombre des atomes du carbone dans une molécule organique).

3. Le théorème 5.1 pour les indices de Harary et l'inverse de Wiener

Nous continuons à vérifier la linéarité de la formule du théorème 5.1 pour certaines valeurs de λ correspondes à des indices particuliers, en se basant sur quelques figures montrant cette corrélation. Par exemple la Figure 5.4, montre la corrélation entre les indices de Harary (correspond à $\lambda = -2$) calculé pour le Kenograph et le Plerograph des isomères de l'octane (c'est-à-dire pour $n = 8 : C_8H_{18}$). On constate que la corrélation est presque linéaire. La Figure 5.5, montre aussi la corrélation entre les indices de l'inverse de Wiener pour Plerograph et Kenograph des mêmes isomères.

Remarque 5.1. On signale juste que les Figures 5.8, 5.4 5.5 et les deux tableaux 5.4 et 5.2 sont tirés tous de l'article [113].

2.2.2 Le théorème 5.1 pour l'indice de Wiener

Parmi les valeurs de λ pour lesquelles la formule du théorème 5.1 est linéaire est celle correspond à l'indice de Wiener. Nous traitons le cas de l'indice de Wiener pour lequel sa valeur calculée sur

5.2 La relation entre les types Kenograph & Plerograph

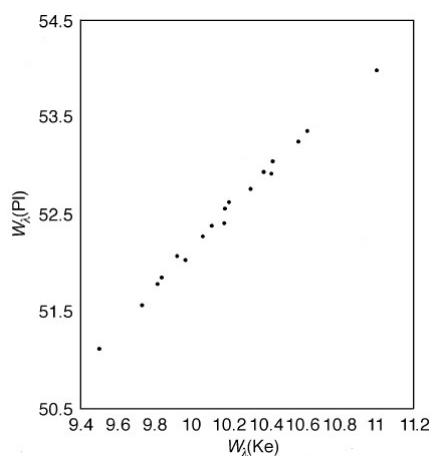


Figure 5.4 – La corrélation entre les indices de Harary du *Pl* et *Ke*

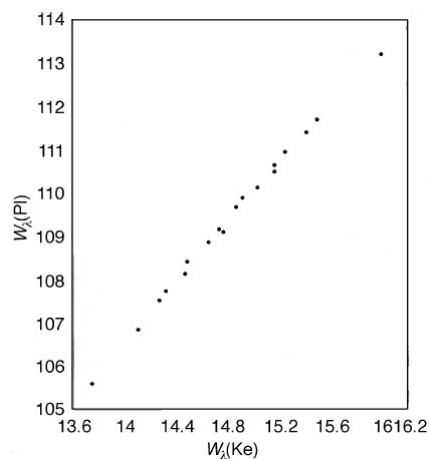


Figure 5.5 – La corrélation entre les indices de l'inverse de Wiener de *Pl* et *Ke*

Kenograph est fortement corrélée et d'une manière linéaire avec celle calculée sur le Plerograph. Le théorème suivant montre cette linéarité.

Théorème 5.2. [113] Soient les représentations Plerograph noté *Pl* et Kenograph noté *Ke* d'une molécule organique de type Alkane. Alors, l'indice de Wiener de ces représentations est liées de la manière suivante :

$$W(Pl) = 9W(Ke) + (3n + 1)^2 \quad (5.3)$$

Démonstration. La démonstration est plus ou moins évidente, elle se base sur la règle de la chimie organiques disant que lorsqu'on veut passer de la représentation Kenograph au Plerograph on ajoute $2n + 2$ feuilles (chaque atome du carbone est lié à quatre atomes d'hydrogène), pour

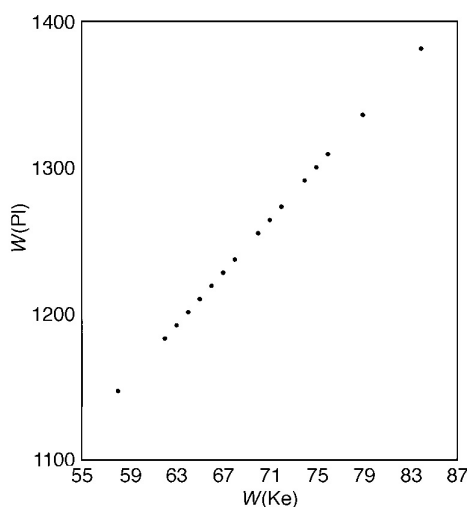


Figure 5.6 – Les indices de Wiener de Ke et Pl pour les isomères de l’Octane

plus de détail voir [113]. □

Pour clarifier cette corrélation nous donnons la Figure 5.6, qui montre la linéarité de cette corrélation. La relation entre $W(Ke)$ et $W(Pl)$ des isomères d’Alcane des molécules d’Octane(C_8H_{18}) est montrée dans cette figure au but de rendre la comparaison de la relation entre $W_1(Pl)$ et $W_1(Ke)$ avec les autres relations approximatives $W_\lambda(Pl)$ et $W_\lambda(Ke)$ pour $\lambda \neq 1$ que nous allons voir par la suite plus claire. Peut-on étendre l’équation 5.3 aux graphes cycliques? c’est

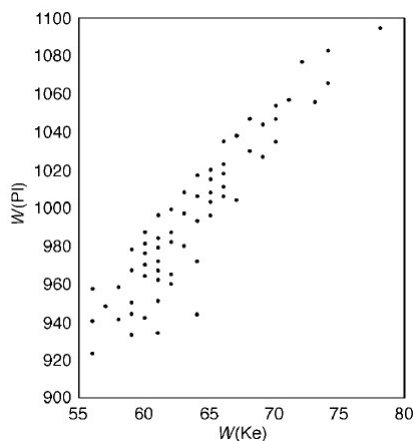


Figure 5.7 – $W(Ke)$ et $W(Pl)$ pour les isomères de Monocyclooctanes

exactement la question posée en [113], et sa réponse est donnée dans la Figure 5.7, qui montre la relation entre les indices de Wiener de Kenograph $W(Ke)$ et de Plerograph $W(Pl)$ pour que les isomères des molécules produisent des graphes qui ne sont pas des arbres, en prenant

5.2 La relation entre les types Kenograph & Plerograph

«Monocyclooctane (C_8H_{16})» comme exemple. Cette figure démontre la complexité de cette relation, qui ne sera jamais linéaire.

2.2.3 La relation entre $W(Pl)$, $W(Ke)$, $TW(Pl)$ et $TW(Ke)$

Nous traiterons dans ce petit paragraph, l'une parmi les plus importantes raisons pour lesquelles le Kenograph est largement utilisé chez la communauté chimiste depuis 1874. Nous allons démontrer que l'indice de Terminal de Wiener du Plerograph $TW(Pl)$ d'un Alkane est en corrélation linéaire exacte avec l'indice de Wiener de Kenograph $W(Ke)$. Donc, les deux indices possèdent précisément la même dépendance structurelle. Mais avant de passer au théorème fondamental de ce paragraph, on va se mettre d'accord sur quelques notations utilisées par la suite, ainsi que quelques résultats que nous aurons besoin pour la démonstration de résultat fondamental. On rappelle d'abord que si on a une représentation Kenograph de n sommets,

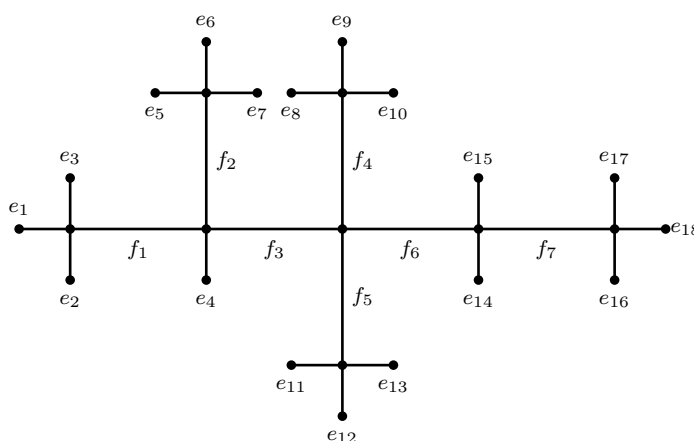


Figure 5.8 – Exemple de nouvelle notation

automatiquement la représentation Plerograph correspondre aura $3n + 2$ sommets, dont $2n + 2$ sont des feuilles. Par la suite on note les arêtes de Kenograph (qui sont les arêtes internes de Plerograph) par f_i et les autres arêtes de Plerograph par e_i (voir la Figure 5.8).

Lemme 5.1. Soient les représentations Plerograph noté Pl , Kenograph noté Ke d'une molécule organique de type Alkane, et $n_i(f|Pl)$ ($p_i(f|Pl)$) le nombre de sommets (de feuilles) du graphe correspond au type Plerograph Pl liant le sommet f des côtés gauche pour $i = 1$ et droit pour $i = 2$. Alors, on a le résultat suivant :

$$n_i(f|Pl) = \frac{3p_i(f|Pl) - 1}{2} \quad (5.4)$$

pour $i = 1, 2$

Démonstration. La démonstration est comme suit :

$$\begin{aligned} n_i(f|Pl) &= 3n_i(f|Ke) + 1 \\ n_i(f|Ke) + p_1(f|Pl) &= 3n_i(f|Ke) + 1 \\ p_i(f|Pl) &= 2n_i(f|Ke) + 1 \\ p_i(f|Pl) &= 2(n_i(f|Pl) - p_1(f|Pl)) + 1 \\ 2n_i(f|Pl) &= 3p_i(f|Pl) - 1. \end{aligned}$$

□

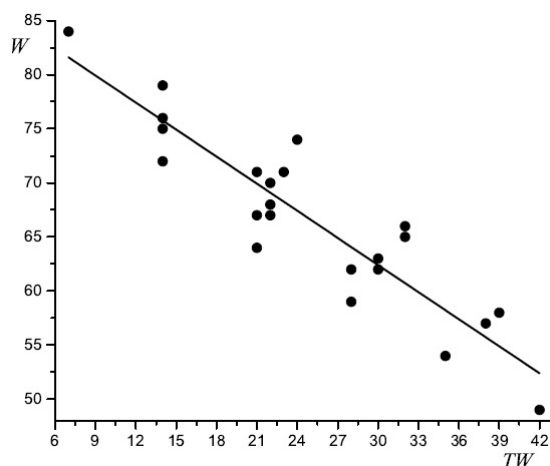


Figure 5.9 – La relation entre W et TW pour les arbres d'ordre 8

On passe maintenant à un résultat intermédiaire et important aussi de point de vu de sujet qu'on est en train de le traiter.

Théorème 5.3. *Soit la représentation Plerograph Pl correspond à une molécule de type Alkane de n atomes du carbonnes. Alors, les indices de Wiener et de Terminal de Wiener de ce type est :*

$$W(Pl) = \frac{1}{4}(9TW(Pl) - 18n^2 - 21n - 5)$$

Démonstration. On utilisant la définition de l'indice de Wiener pour les arbres et le lemme 5.1,

5.2 La relation entre les types Kenograph & Plerograph

on a :

$$\begin{aligned}
 W(Pl) &= \sum_v n_1(v)n_2(v) \\
 &= \sum_f n_1(f)n_2(f) + \sum_e n_1(e)n_2(e) \\
 &= \sum_f \left(\frac{3p_1(f|Pl) - 1}{2}\right) \left(\frac{3p_2(f|Pl) - 1}{2}\right) + (2n + 2)(3n + 1) \\
 &= \frac{9}{4} \sum_f (p_1(f|Pl)p_2(f|Pl)) - \frac{3}{4} \sum_f k + \frac{1}{4} \sum_f (1) + (2n + 2)(3n + 1) \\
 &= \frac{9}{4} TW(Pl) + (2n + 2) \left(-\frac{9}{4}(2n + 1) - \frac{3}{4}(n - 1) + 3n + 1\right) + \frac{1}{4}(n - 1) \\
 &= \frac{9}{4} TW(Pl) - \frac{1}{4}(18n^2 + 21n - 5).
 \end{aligned}$$

□

On passe maintenant au résultat principal dans ce paragraphe.

Théorème 5.4. *Soient les représentations Plerograph noté Pl et Kenograph noté Ke d'une molécule organique de type Alkane. Alors, les indices de Wiener et de Terminal de Wiener de ces deux représentations peuvent s'exprimer comme suit :*

$$TW(Pl) = 4W(Ke) + 6n^2 + 5n + 1 \tag{5.5}$$

Démonstration. On utilise les théorèmes 5.2 et 5.3, on a :

$$\begin{aligned}
 9W(Ke) + (3n + 1)^2 &= \frac{1}{4}(9TW(Pl) - 18n^2 - 21n - 5) \\
 TW(Pl) &= \frac{4}{9}(9W(Ke) + (3n + 2)^2 + \frac{9}{2}n^2 + \frac{21}{4}n + \frac{5}{4}).
 \end{aligned}$$

□

Nous avons vu donc, à travers cet important résultat (le théorème 5.4) et à travers aussi la Figure 5.9, que dans le cas des Alkanes, non seulement l'indice de Wiener de la représentation Plerograph a précisément une dépendance structurelle comme l'indice de Wiener de celle de Kenograph (le théorème 5.2), mais ça reste aussi valable pour l'indice de Terminal de Wiener de la représentation Plerograph $TW(Pl)$. Et donc, par conséquence si les indices de Terminal de Wiener et de Wiener sont pris en considération dans les méthodes de classifications, on n'aura pas besoin du Plerograph et on peut travailler juste par le Kenograph pour faciliter les calculs.

2.3 La corrélation entre les TI basés sur le degré

L'utilisation des descripteurs de la structure moléculaire représentée par des graphes est devenue obligatoire. Surtout dans le processus de la découverte de médicaments. Ces descripteurs sont basés en leurs définitions sur plusieurs matières premières du graphe comme la distance, les degrés des sommets ou bien les deux. On continue dans ce petit paragraphe la découverte des relations qui lient les deux représentations d'une molécule : Kenograph et Plerograph. Dans le précédent paragraphe nous avons étudié les indices basés sur la distance en leurs définitions, et nous avons vu comment ces indices relient les deux types de représentations moléculaires. Dans ce paragraphe on étudiera les indices basés sur le degré en leurs définitions et on étudiera aussi leurs influences sur la relation entre les types Kenograph et Plerograph.

2.3.1 La formule de base

Un très grand nombre d'indices topologiques sont basés, en leurs définitions, sur les degrés des sommets d'un graphe. Donc, on peut leur donner cette définition générale.

Théorème 5.5. *Soit G un graphe simple, planaire et connexe. Donc, tout indice topologique TI basé en sa définition sur les degrés des sommets de G peut s'écrire comme suit :*

$$TI(G) = \sum_{1 \leq i \leq j \leq n-1} F(i, j) m_{ij} \quad (5.6)$$

Où m_{ij} est le nombre des arêtes de G connectées aux sommets de degrés i et j respectivement.

Démonstration. En générale ce théorème, n'est que le résultat d'une remarque approfondie des différentes définitions des indices utilisant en leurs définitions le degré d'un sommet du graphe. Car tout indice basé sur le degré peut s'écrire comme suit :

$$TI(G) = \sum_{e \in E(G)} F(deg(u), deg(v)). \quad (5.7)$$

Avec $deg(u)$ et $deg(v)$ sont les degrés respectivement des sommets construits l'arête e du graphe G . □

La plupart des définitions des indices topologiques actuels, définis par les degrés, correspondents à la fonction $F(x, y)$. Par exemple les définitions des indices de Randić [154][33], de la bonde-atom de connectivité [65][144], de Zagreb augmenté [30, 201] et d'Harmonique [146] correspondants à la fonction $F(x, y)$ sont respectivement $\frac{1}{\sqrt{xy}}$, $\sqrt{\frac{(x+y-2)}{xy}}$, $[\frac{xy}{(x+y-2)}]^3$ et $\frac{2}{(x+y)}$.

2.3.2 Le cas de Plerograph

Lors qu'on voit la Figure 1.5, représentant un exemple de Kenograph et de Plerograph des Hydrocarbures, on constate que toute représentation Plerograph n'a que des sommets u de degré soit $deg(u) = 1$ soit $deg(u) = 4$, ce qui exige que leurs arêtes ne sont que des arêtes ayant des extrémités de 1 ou 4. Et donc la formule 5.6 pour le Plerograph devient.

Corollaire 5.1. *Les indices topologiques basés sur les degrés du graphe de Plerograph sont définis comme suit :*

$$TI(Pl) = hF(1, 4) + (2n - \frac{h}{2}F(4, 4)) \quad (5.8)$$

Où $h = m_{14}$

Démonstration. Puisqu'on a dit que les arêtes du Plerograph ne possède que les extrémités des degrés 1 et 4 ou de 4 et 4, l'équation 5.6 devient :

$$TI(Pl) = F(1, 4)m_{14} + F(4, 4)m_{44}. \quad (5.9)$$

En considérant la formule saturée d'hydrocarbure C_nH_{2n+2} et en posant $h = m_{14} = 2n + 2$ et $4n + h = 2(m_{14} + m_{44})$ on parvient notre formule. \square

Ce corollaire, implique que tout indice topologique basé sur les degrés des sommets des graphes Plerograph ne dépend que de nombre des atomes du carbone et d'hydrogène de la molécule correspondante. Ce qui veut dire aussi que tous les isomères d'un même nombre de sommets auront la même valeur de $TI(Pl)$. Et ça va expliquer que les indices topologiques basés sur les degrés de Plerograph sont complètement inutiles comme descripteurs de structure, car ils ne font pas la différence entre les isomères de la même molécule.

2.3.3 L'étude de quelques TI basé sur les valeurs propres

Ce paragraphe va se consacrer à la comparaison de la corrélation des indices de spectral radius (égale à plus grande valeur propre, voir [55, 112] pour plus de détails sur cet indice) et d'Estrada (égale à la somme de $\exp(\lambda_i)$ sur toutes les valeurs propres du graphe correspondu, voir [64, 66?]] pour plus de détails sur cet indice) pour les deux graphes Plerograph et Kenograph. Ces deux indices sont du type des indices basés sur les valeurs propres de la matrice du graphe représentant la molécule en question.

La figure 5.10 montre la très bonne corrélation entre les indices de Spectral radius du Kenograph $\lambda_1(Ke)$ et du Plerograph $\lambda_1(Pl)$, pour les isomères d'une molécule appelée Undecanes de la formule $C_{12}H_{26}$. Ce qui a donné 355 données (c'est-à-dire 355 molécules de 12 atomes du

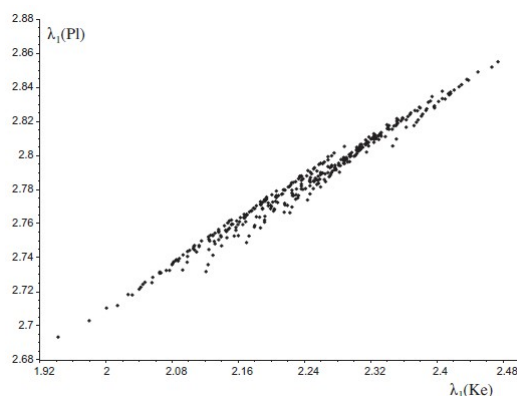


Figure 5.10 – La corrélation entre $\lambda_1(Ke)$ et $\lambda_1(Pl)$ pour les isomères d'Undecanes ($C_{12}H_{26}$)

carbone). Cette corrélation a comme coefficient 0.992 ce qui veut dire que presque 1. Donc, la situation avec les descripteurs basés sur les valeurs propres n'est pas très bonne, par ce que cette corrélation est assez bonne que la dépendance structurelle des deux types, est pratiquement identique. Le cas de l'indice d'Estrada reste différent, comme dans la Figure 5.11, qui

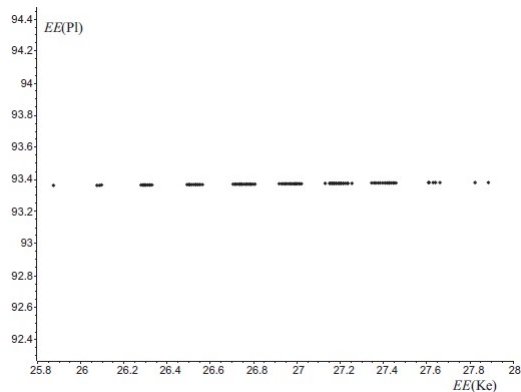


Figure 5.11 – La corrélation entre $EE(Ke)$ et $EE(Pl)$ pour les isomères d'Undecanes ($C_{12}H_{26}$)

montre les indices d'Estrada des Plerograph $EE(Pl)$ et du Kenograph $EE(Ke)$ pour les isomères d'Undecane. La figure montre que l'indice d'Estrada de Plerograph $EE(Pl)$ est presque indépendant de la structure moléculaire, ce qui n'est pas le cas avec l'indice d'Estrada de Kenograph $EE(Ke)$ dont les valeurs varient dans un intervalle assez grand. Ce qui explique que les deux représentations sont totalement indépendantes dans le cas l'indice d'Estrada.

3 La dépendance structurelle de $TW(G)$

Vu l'existence d'une relation linéaire exacte entre les indices de Wiener calculés en utilisant les représentations Kenograph $W(Ke)$ et Plerograph $W(Pl)$ des isomères d'Alkane (voir les précédents paragraphes) et puisque l'indice de Wiener est le prédécesseur de l'indice de Terminal de Wiener on a essayé dans le papier [108] de décortiquer la relation entre les indices de Terminal de Wiener calculés en utilisant les représentations Kenograph $TW(Ke)$ et Plerograph $TW(Pl)$ croyant qu'on pouvait les relier. Après plusieurs efforts on a su à la fin que cette relation ne sera jamais linéaire ni approximativement linéaire. Mais, on a découvert que elle possède un comportement particulier (voir la Figure 5.12) nécessite une étude particulière. Car nous avons constaté que l'indice de Terminal de Wiener TW dépend de la structure des molécules. Alors, cette section va étudier la dépendance de l'indice de Terminal de Wiener aux structures Kenograph Ke et Plerograph Pl .

3.1 Problématique & discussion

3.1.1 Problématique

En voyant la corrélation entre l'indice de Wiener W et celui de Terminal de Wiener TW (voir la Figure 5.9 et les précédents paragraphes), la relation linéaire exacte entre l'indice de Wiener de la représentation Plerograph $W(Pl)$ et celui de Wiener de la représentation Kenograph $W(Ke)$ qu'on a vu dans l'article [113] et la relation entre l'indice de Terminal de Wiener de Plerograph $TW(Pl)$ et l'indice de Wiener de Kenograph $W(Ke)$ (voir le paragraphe précédent) en prévoyant qu'aussi l'indice de Terminal de Wiener de Plerograph $TW(Pl)$ et celui de Kenograph $TW(Ke)$ sont linéairement corrélés ou au moins juste corrélés. Mais, lorsqu'on constate attentivement la Figure 5.12 on se surpris de la forme de cette relation. Cette dernière qui se constitue de 5 groupes (pour $p = 2, 3, 4, 5, 6$, de gauche à droite) de points de données (chaque point de données correspond à un isomère, c'est-à-dire une molécule) et cela juste pour les isomères des Nonanes, c'est-à-dire les molécules de $n = 9$ atomes du carbone, et chaque groupe correspond à un nombre de feuilles p . La Figure 5.12 montre la corrélation entre les indices de Terminal de Wiener de Kenograph $TW(Ke)$ et Plerograph $TW(Pl)$ appliquée sur les isomères de Nonanes, c'est-à-dire les molécules de la formule C_9H_{20} . Le problème qui demande qu'il faut prendre en considération est que, lorsqu'on aura $n \leq 9$ ces groupes sont disjoints, comme il montre la Figure 5.12. Mais dès qu'on dépasse 9 sommets les groupes peuvent être disjoints comme ils peuvent s'intervenir entre eux. Ce qui veut dire que cette relation demande une étude approfondie avec des explications bien détaillées. Notre problématique dans cette

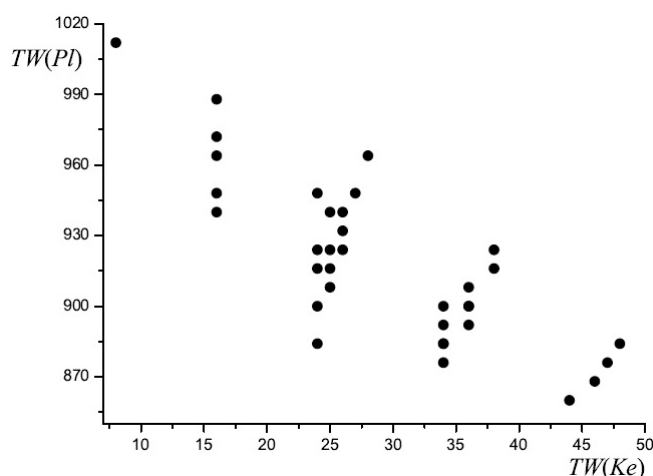


Figure 5.12 – La corrélation entre $TW(Ke)$ et $TW(Pl)$ pour les isomères de Nonanes

section est : **A quel ordre la relation entre $TW(Ke)$ et $TW(Pl)$ diffère de celle de $W(Pl)$ et $W(Ke)$?**

Avant de passer à nos principaux résultats de cette étude, on va tout d'abord discuter les groupes de cette relation et comment on peut les relier à de tels ou tels p nombre de feuilles de chaque isomère, soit de Kenograph Ke soit de Plerograph Pl . Et à partir de cette discussion nous pourrions arriver à comprendre le résultat qui pourrait s'apparaître un peu compliqué à la compréhension. Pour faciliter la discussion d'une part et simplifier l'écriture d'autre part, nous devons nous mettons d'accord sur quelques notations ainsi que quelques données de base. On dit qu'une arête d'un arbre T est de (p_1, p_2) - type, si ses deux côtés ont p_1 et p_2 feuilles. Par la suite, un Kenograph noté Ke est supposé posséder n sommets, c'est-à-dire qu'il représente un Alkane C_nH_{2n+2} , où n est le nombre de carbone de cet Alkane. Et la représentation Plerograph correspond à cet Alkane possède $3n+2$ sommets dont $2n+2$ sont des feuilles. Tout Kenograph Ke d'ordre n possède $n-1$ arêtes. On note par la suite le nombre de feuilles de chaque Kenograph Ke par p .

3.1.2 Discussion

Puisque la définition de l'indice de Terminal de Wiener se base sur le produit du nombre des feuilles existantes aux extrémités de chaque arête (p_1/p_2) du graphe (voir la définition 4.2) les groupes construits aux figures des isomères, comme la Figure 5.12, s'influencent automatiquement au nombre de feuilles noté par p dans chaque isomère. Lors de l'inspection de la Figure 5.12 ou les autres figures pour $n \neq 9$ on déduit automatiquement que les points de données sont regroupés en quelques groupes disjoints. Et d'après ce qu'on a dit au début, il n'est pas

difficile de savoir que chacun de ces groupes est déterminé par une valeur particulière de p , que nous allons les discuter une par une.

1. **Pour un nombre n petit de sommets**

– **Pour $p = 2$**

Lorsqu'on voit les isomères de toute molécule, on est sûr qu'il existe un seul Kenograph d'ordre n où $p = 2$, et qui est tout simplement le graphe Path (voir la figure 3.7) correspondant à un Alkane normal. Donc, c'est facile de le localiser dans la Figure 5.12, et qu'il correspond au seul point de données se trouvant au côté le plus gauche de cette figure. Puisque toutes les arêtes de tel Kenograph sont de $(1, 1)$ – *type* et puisque la taille de tout Path est $n - 1$ donc, l'indice de Terminal de Wiener d'unique Kenograph de $p = 2$ et $n \geq 2$ est :

$$TW(Ke) = 1 \times 1(n - 1) = (n - 1). \quad (5.10)$$

– **Pour $p = 3$**

Si dans la représentation Kenograph le nombre de feuilles $p = 3$ alors, toute arête de ce graphe est de $(1, 2)$ – *type*. Et puisque tous les isomères de tel Kenograph n'ont que ces types d'arêtes l'indice de Terminal de Wiener de tout isomère de ce Kenograph pour $p = 3$ et pour $n \geq 4$ sera :

$$TW(Ke) = 1 \times 2(n - 1) = 2(n - 1). \quad (5.11)$$

Car on a $n - 1$ arêtes.

Cette équation signifie que tous les isomères d'Alkane ayant qu'un seul atome du carbone du degré 3 (c'est-à-dire lie à trois autres atomes du carbone) du Kenograph ont la même valeur de Terminal de Wiener $TW(Ke)$, par contre chaque isomère possède sa propre valeur de Terminal de Wiener de Plerograph $TW(Pl)$. Et si on analyse la Figure 5.12 et ses groupes, on déduit rapidement que le groupe des points de données correspond à ce type des isomères ($p = 3$) n'est que le deuxième groupe du côté le plus à gauche (La seule ligne verticale. Et si on remplace dans l'équation 5.3 n par 9 on trouve exactement $TW(Ke) = 16$ pour tout le groupe et différentes valeurs de Plerograph $TW(Pl)$).

– **Pour $p = 4$**

Dans le cas où $p = 4$, on aura deux combinaisons, soit $(1, 3)$ soit $(2, 2)$. C'est-à-dire que toute arête de Kenograph est : soit de $(1, 3)$ – *type* ou bien de $(2, 2)$ – *type*. La

deuxième chose attirante de ce cas est que l'isomère du plus petite valeur de Terminal de Wiener $TW(Ke)$ est celui qui possède au moins 4 arêtes de $(1, 3) - type$. Et celui de plus grande valeur de Terminal de Wiener $TW(Ke)$ est celui possédant au plus 4 arêtes de $(1, 3) - type$ et toutes les autres sont de $(2, 2) - type$. Donc, tous les isomères de Kenograph où $p = 4$ et $n \geq 5$ ont comme indice de Terminal de Wiener pour le Kenograph $TW(Ke)$:

$$1 \times 3(n - 1) = 3n - 3 \leq TW(Ke) \leq 1 \times 3(4) + 2 \times 2(n - 5) = 4n - 8. \quad (5.12)$$

Et comme résultat reservoir, avant de le démontre dans la sous-section suivante, si on remplace dans l'inégalité 5.12 n par 9 on tombe au troisième groupe du plus gauche de la Figure 5.12. Ce qui veut dire que ce groupe corresponde au cas où $p = 4$.

– **Pour** $p = 5$

Maintenant, pour les isomères où $p = 5$ les arêtes sont, soit de $(1, 4) - type$ soit de $(2, 3) - type$, et l'isomère de plus petite valeur de Terminal de Wiener $TW(Ke)$ est celui qui possède au moins 5 arêtes de $(1, 4) - type$ et au moins une de $(2, 3) - type$. Et l'isomère qui aura la plus grande valeur de Terminal de Wiener $TW(Ke)$ est celui ayant au plus 5 arêtes de $(1, 4) - type$ et toutes les autres de $(2, 3) - type$. Pour tout isomère de 5 feuilles et $n \geq 7$, on a :

$$1 \times 4(n - 2) + 2 \times 3(1) = 4n - 2 \leq TW(Ke) \leq 1 \times 4(5) + 2 \times 3(n - 6) = 6n - 16. \quad (5.13)$$

Si on remplace aussi dans l'inégalité 5.13 n par 9 on remarque que les isomères correspondent au quatrième groupe du plus gauche.

– **Pour** $p \geq 5$

Si on suit la même méthodologie pour les autres cas où $p \geq 5$ on aura des groupes différents pour toute valeur de p . par exemple pour $p = 6$ et $n \geq 8$ on aura :

$$5n - 1 \leq TW(Ke) \leq 9n. \quad (5.14)$$

Et pour $p = 7$ et $n \geq 10$ on aura :

$$6n + 6 \leq TW(Ke) \leq 12n - 54. \quad (5.15)$$

Et cette situation nous confirme l'hypothèse qu'on a déjà supposé, disant que chaque groupe de points de données correspond à une valeur particulière de $TW(Ke)$.

2. Pour un nombre n grand de sommets

Lorsqu'on analyse les équations 5.1-5.15, on constate que les groupes correspondants sont déjoints et ça pour $n \leq 10$ et $p \leq 5$. Mais, dès qu'on augmente la valeur du nombre des atomes du carbone le nombre de feuilles donne plusieurs combinaisons, dès que les groupes s'interferent et perturbe les groupes. Pour clarifier ces remarques, on donne quatre

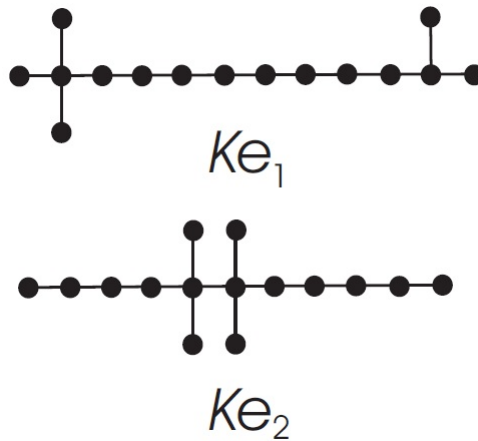


Figure 5.13 – 2 Kenographs de $n = 15$ avec des p différentes et de même $TW(Ke)$

exemples, qu'on a traité dans le même article [108]. Pour $n = 15$ par exemple, il existe des Kenographs avec deux cas de nombre de feuilles différents $p = 5$ et $p = 6$ (voir la Figure 5.13) mais les deux isomères possèdent la même valeur de Terminal de Wiener $TW(Ke)$. Et pour $n \geq 15$ quelques Kenographs avec $p = 5$ ont une valeur de Terminal de Wiener plus grande que quelques Kenographs avec $p = 6$. Maintenant, si on prend par

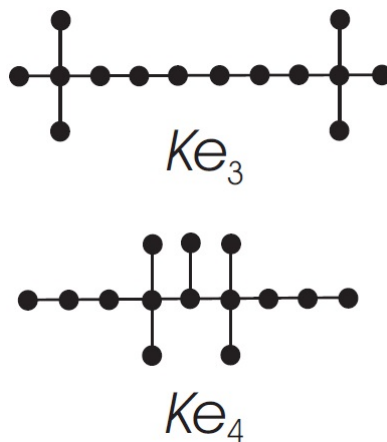


Figure 5.14 – 2 Kenographs de $n = 14$ avec des $p_1 \leq p_2$ mais $TW(Ke_{p_1}) \geq TW(Ke_{p_2})$

exemple $n = 13$ on va trouver des Kenographs avec $p = 6$ et $p = 7$ ont la même valeur de Terminal de Wiener. Et si on considère $n \geq 13$ on trouve des Kenographs avec $p = 6$ ont une valeur de Terminal de Wiener plus grande que quelques Kenographs où leurs p est égale à 7 (voir par exemple la Figure 5.14).

Après la précédente discussion nous avons arrivé à comprendre le comportement de la relation entre les Terminal de Wiener de Kenograph $TW(Ke)$ et de Plerograph $TW(Pl)$. La suite de ce chapitre discutera deux résultats intéressants qui peuvent être utiles aux chimistes qui cherchent l'explication du comportement de cette relation, un pour Kenograph et l'autre pour Plerograph.

3.2 La dépendance de l'indice de Terminal de Wiener au Kenograph

Pour faciliter le passage au cas générale de l'indice de Terminal de Wiener pour le type Kenograph, on était obligé de revenir au cas de $p = 3$ où on a une seule valeur pour tous les isomères de type Kenograph et des valeurs différentes pour chaque isomère de type Plerograph. Puis le comparer au cas de $p \geq 4$ où on a plusieurs sous-groupes de chaque cas de p . Ce qui nous a demandé de chercher la cause de ces différents comportements de $TW(Ke)$ d'un cas à un autre. A la fin on est arrivé à déterminer la cause derrière ces comportements. On a compris que ces différentes valeurs lors du passage du cas où $p = 3$ à celles du cas où $p \geq 4$ venu du faite que dans le premier cas on a qu'une seule permutation, c'est-à-dire on a que les arêtes de $(1, 2) - type$. Par contre le cas par exemple de $p = 4$, on a deux permutations, c'est-à-dire on a les arêtes de $(1, 3) - type$ et celles de $(2, 2) - type$, et ainsi de suite pour les autres cas où $p \geq 5$.

3.2.1 Le résultat principal

On a été dirigé à l'astuce suivant, on change la place des arêtes et donc automatiquement son type sera changé et tout ça va influencer d'une manière indirecte aux valeurs de $TW(Ke)$. Notre astuce est expliqué à la Figure 5.15. Dans ce schéma on explique le passage du Kenograph Ke à sa transformation Ke' . Dans le Kenograph Ke les arêtes uv , ij et jk sont respectivement de $(q, p - q) - type$ (q feuilles au côté droit de l'arête uv et $p - q$ feuilles dans l'autre côté), $(1, p - 1) - type$ et $(1, p - 1) - type$. Et dans sa transformation Ke' on a changé la place d'une arête du faite qu'il devient les arêtes uj , ju et ik respectivement de $(q, p - q) - type$, $(q, p - q) - type$ et $(1, p - 1) - type$ où q est entier inférieur ou égale à p (voir la figure 5.15). Le théorème suivant nous explique ce qui se passe lorsqu'on change le $(p_1, p_2) - type$ d'une seule arête.

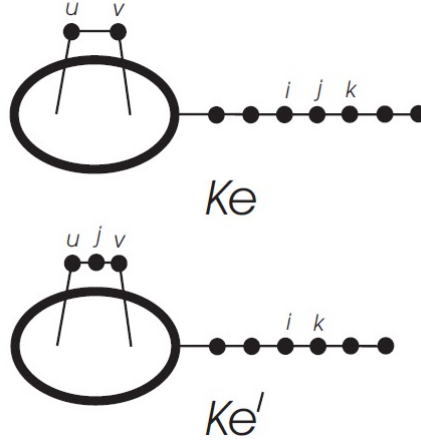


Figure 5.15 – Le schéma explicatif de transformation Ke à Ke'

Théorème 5.6. Soit la représentation Kenograph Ke possédant p feuilles, et dont ses détails sont expliqués dans la Figure 5.12. Faisant la transformation de Ke à Ke' , comme il montre la même figure. Et soit aussi l'arête uv de la même représentation Ke de $(q, p - q)$ - type, où $q \leq p$. Alors :

$$TW(Ke') - TW(Ke) = q(p - q) - (p - 1). \quad (5.16)$$

Démonstration. D'après ce qu'on a expliqué à la Figure 5.15, et en appliquant la définition de Terminal de Wiener, on aura :

$$\begin{aligned} TW(Ke) &= q(p - q) + (p - 1) + (p - 1) + TC \\ TW(Ke') &= q(p - q) + q(p - q) + (p - 1) + TC \end{aligned}$$

Où TC est la partie de l'équation commune entre Ke et Ke' . En faisant la différence entre les deux équations on parvient à l'équation 5.16. \square

Remarque 5.2. Si $q = 1$, nous trouvons le cas trivial où $TW(Ke') = TW(Ke)$, et si $q \geq 2$, on aura $TW(Ke') > TW(Ke)$.

3.2.2 Discussion

Pour enrichir les discussions faites au début de cette section, nous avons appliqué le précédent théorème aux cas de la Figure 5.12. Ces cas sont cités dans le corollaire suivant :

Corollaire 5.2. Les cas spéciales suivant sont très importants :

1. Si $p = 4$, et puisque $q \leq |p - 2| = 2$, donc obligatoirement $q = 2$, et par conséquence $TW(Ke') - TW(Ke) = 1$.

2. Si $p = 5$, et puisque $q \leq |p - 2| = 2$, donc obligatoirement $q = 2$, et par conséquence $TW(Ket) - TW(Ke) = 2$.
3. Si $p = 6$, et puisqu'on a soit $q = 2$ ou bien $q = 3$, on aura par conséquence soit $TW(Ket) - TW(Ke) = 3$ ou bien $TW(Ket) - TW(Ke) = 4$.

Comme on a dit, ce corollaire nous explique les comportements anormaux de la Figure 5.12. Donc, à partir de cette figure, nous constatons que la distance entre les lignes verticales dans le cas où $p = 4$ est égale à 1; et c'est exactement ce qu'on a trouvé au corollaire 5.2. Et la distance entre celles du cas $p = 5$ on constate aussi qu'est égale à 2, ce qui confirme le même corollaire. Ce dernier nous garantie une explication similaire aux cas où $p \geq 6$, mais avec plus de complexité.

3.3 La dépendance de l'indice de Terminal de Wiener au Plerograph

L'étude de la dépendance des valeurs de $TW(Pl)$ à la structure moléculaire est plus compliquée si l'on compare avec la dépendance de celles de $TW(Ke)$. Car, comme on le sait depuis le premier chapitre de ce mémoire, Le type Plerograph comporte plus de sommets que son analogue Kenograph. Et autant que les sommets s'ajoutent autant que le nombre des isomères s'augmente, ce que veut dire d'une façon ou d'une autre que les arêtes de $(p_1, p_2) - type$ se multiplient. Et on sera à la fin dans une situation plus compliquée, et difficile ne serait que pour l'analyser. Afin de montrer la complexité de cette situation on a étudié dans le même article des nos précédents travaux, juste le cas de Plerograph dont $p = 3$. Et comme on a vu précédemment, ce cas est le plus simple, et où tous les Kenograph ont la même valeur de TW .

3.3.1 Le résultat principal

Notons par $Ke(a_1, a_2, a_3)$ le Kenograph d'ordre n possédant exactement un seul sommet de degré 3, avec lequel est attaché a_1 sommets d'un côté, a_2 sommets d'autre côté et a_3 sommets d'autres côté. C'est-à-dire qu'on a $n = a_1 + a_2 + a_3 + 1$. Supposant que $a_1 \leq a_2 \leq a_3$. On note le Plerograph correspond à ce Kenograph par $Pl(a_1, a_2, a_3)$ (voir la Figure 5.16). On note aussi par $p^* = 2n + 2$ le nombre de feuilles de $Pl(a_1, a_2, a_3)$, et donc automatiquement ce Plerograph aura p^* arêtes de $(1, p^* - 1) - type$. Le théorème suivant donne la valeur de Terminal de Wiener de Plerograph $TW(Pl)$.

Théorème 5.7. Soient $n \geq 4$, et $a_1 + a_2 + a_3 = n - 1$ avec $a_1 \leq a_2 \leq a_3$. Donc,

$$TW(Pl(a_1, a_2, a_3)) = 2(n - 1) \sum_{k=1}^3 a_k^2 - \frac{4}{3} \sum_{k=1}^3 a_k^3 + \frac{7}{3}n + \frac{5}{3}. \quad (5.17)$$

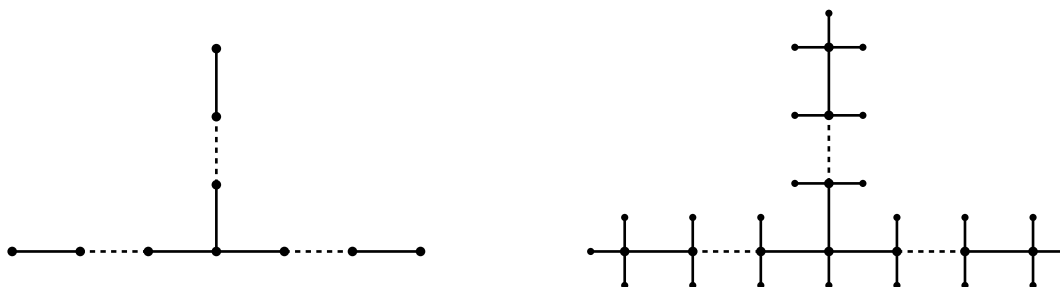


Figure 5.16 – Exemple du Kenograph $Ke(a_1, a_2, a_3)$ et son Plerograph

Démonstration. Comme on a dit, le Plerograph $Pl(a_1, a_2, a_3)$ a $p^* = 2n + 2$ feuilles est donc p^* arêtes de $(1, p^*)$ – type. En appliquant la définition de l'indice de Terminal de Wiener, on a :

$$\begin{aligned} TW(Pl(a_1, a_2, a_3)) &= p^*[(p^* - 1) \times 1] + \sum_{k=1}^3 \sum_{i=1}^{a_k} (2i + 1)[p^* - (2i + 1)] \\ &= (2n + 2)(2n + 1) + \sum_{k=1}^3 \sum_{i=1}^{a_k} (2i + 1)(2n - 2i + 1) \\ &= 2(n - 1) \sum_{k=1}^3 a_k^2 - \frac{4}{3} \sum_{k=1}^3 a_k^3 + \frac{7}{3}n + \frac{5}{3}. \end{aligned}$$

□

3.3.2 Discussion

Donc, si on constate profondément le théorème 5.7, on déduit facilement que l'indice de Terminal de Wiener du Plerograph $TW(Pl)$ dépend de la longueur des trois branches, autrement dit, il dépend de a_1 , a_2 et a_3 . Et en se basant sur le théorème 5.7, Le théorème suivant, nous donnerons le Plerograph de plus grande et plus petite valeur.

Théorème 5.8. Soient $n \geq 4$, et $a_1 + a_2 + a_3 = n - 1$ avec $a_1 \leq a_2 \leq a_3$. Donc,

- La valeur de $TW(Pl(a_1, a_2, a_3))$ est la plus grande si et seulement si $a_1 = a_2 = 1$ et $a_3 = n - 3$.
- La valeur de $TW(Pl(a_1, a_2, a_3))$ est la plus petite si et seulement si $a_3 - a_1 \leq 1$.

Démonstration. Evident.

Dans ce chapitre, on a analysé les différentes relations et/ou corrélation reliant les structures des molécules aux expressions des indices topologiques en particulier les indices de Wiener et celui de Terminal de Wiener. Et cette étude est très importante au moins pour les chimistes qui cherchent toujours à comprendre les comportements inhabités de leurs molécules, surtout celles de la chimie organique comme les Alkanes par exemple. Mais si on élargie le cercle de cette étude,

Chapitre 5. Les représentations Kenograph et de Plerograph

on trouve qu'automatiquement elle atteint d'une manière ou d'une autre les biologistes aussi. Car tous les protéines ne sont que des molécules. Maintenant qu'on a compris l'importance de ce travail au niveau de la chimie et la biologie, on sera capable de comprendre son rôle radicale pour le processus de la découverte de nouveaux médicaments ou ce qu'on appelle par «Drag discover» que nous avons déjà introduit dans nos précédents chapitres.

Conclusion générale et perspectives

Dans le cadre de cette thèse, nous avons étudié quelques indices topologiques ; comme l'indice de Wiener, l'indice d'Hyper-Wiener, l'indice de degré de distance et l'indice de Terminal de Wiener. Les molécules chimiques sont modélisées par des graphes moléculaires très complexes. Le but de notre thèse est de trouver des formules qui permettent de calculer des indices topologiques des graphes moléculaires plus facilement. en améliorant leurs formules de définition, d'une part et de voir les différentes relations qui peuvent être créés entre eux d'autre part, en les regroupant en des relations particulières selon leur matière première de définition (comme la distance entre deux sommets, le degré de sommet, ...) afin d'améliorer les méthodes de sélection de variables qui sont l'une des principaux étapes de processus de la modélisation. Durant ces années de thèse nous avons aussi étudié les différentes façons de transformer la molécule ou le composé moléculaire, en particulier ceux qui sont de type organique, en un graphe planaire simple et connexe qui a été le sujet central de cette thèse en rendant service à nos collègues les chimistes qui ont utilisé la transformation Kenograph pendant cent vingt années sans chercher les raisons de cette domination de cette représentation sur celle de Plerograph. Les techniques traditionnelles de modélisation établissent une relation entre la grandeur modélisée et un vecteur de variables qui la déterminent. Les principaux obstacles de ces méthodes résident dans la difficulté du choix des variables pertinentes lors de la phase de modélisation, et dans leur calcul ou leur mesure préalable. Dans ce rapport qui a été formé de trois principales parties, dont la première est dédiée aux concepts de base de la théorie de graphe (définitions et propriétés) dans son premier chapitre et à l'explication de la notion de prédiction de propriétés et d'activités de molécules à travers les études *QSAR* et *QSPR* en détaillant leurs principaux étapes ainsi que leurs différentes méthodes dans son deuxième chapitre. Le troisième chapitre qui s'intègre dans la deuxième partie de ce rapport a été consacré aux principales méthodes de calcul de l'indice de Wiener soit pour les graphes en générale soit pour les arbres en particulier et le

quatrième chapitre qui est également appartient à cette même partie traite les autres indices topologiques (les indices de Hyper-Wiener, de degré de distance et de Terminal de Wiener) et les leurs méthodes de calcul. La modélisation moléculaire et le criblage virtuel sont des domaines de la chimo-Informatique qui exigent une interaction très forte avec différents métiers afin de relier les résultats obtenus sur ordinateur avec des données expérimentales. De plus, les données structurales fournies par les bases de données, en particulier, de molécules constituent la base de ces méthodologies. Le modélisateur est donc souvent au centre d'un réseau de collaborations qui fait intervenir des mathématiciens, des informaticiens, des biologistes ou des chimistes de synthèse. Au cours de ces trois années de thèse, j'ai eu l'occasion de participer à deux projets de recherche avec l'un des pionniers de la théorie des graphes moléculaires et l'un des ses fondateurs, Monsieur Ivan Gutman. La collaboration avec lui nous a fait découvert la problématique de la représentation moléculaire, et la dispute entre quelques chercheurs en particulier les contemporains sur l'utilité et l'inutilité de la représentation de Plerograph en présence de celle de Kenograph. Dans le dernier chapitre de ce rapport qui compose le seul chapitre de la dernière partie nous avons présenté les plus importants résultats de cette collaboration comme : les différentes corrélations entre quelques indices topologiques calculés sur le Kenograph et sur le Plerograph en raison de prouver l'inutilité de type Plerograph en présence de type Kenograph. En particulier, la corrélation entre l'indice de Wiener de Plerograph $W(Pl)$ et l'indice de Terminal de Wiener de Kenograph $TW(Ke)$ avant que nous les généraliser sur les graphes de Thorn et son graphe de base. Cette corrélation linéaire nous a confirmé la non importance de type Plerograph en présence de Kenograph, ce qui ajoute à la littérature une autre raison explicative du choix de Kenograph sur le Plerograph. Nous avons également analysé la relation entre les indices de terminal de Wiener de Kenograph $TW(Ke)$ et de Plerograph $TW(Pl)$. Cette relation même s'elle n'apparaît pas en claire, nous l'a traité de point de vu nombre de feuilles dans chaque représentation Kenograph et son influence sur la valeur de $TW(Ke)$ et $TW(Pl)$. Ces deux déreniers résultats nous a donné un bagage riche des relations entre différents indices topologiques, ce bagage nous a aidé à regrouper plusieurs indices de mêmes nature, c'est-à-dire leurs définitions se basent sur la même chose (distance, degré, valeurs propres,...). Le regroupement de plusieurs indices en une seule relation sera sans doute l'une des plus importantes techniques de méthodes de la sélection de variables.

Les perspectives :

Afin de poursuivre notre chemin de recherche dans cette discipline nous prévoyons les perspectives suivantes :

- L'applications de nos principales contributions aux différents types de composition (comme, la composition de Corona, du produit d'arête, du produit de sommet, ...).

- Et puisque nos méthodes utilisent $d_G(k)$ qui n'a aucune relation avec la connectivité des graphes et puisque aussi il y a une formule pour calculer l'indice de Wiener des graphes non connexes [107] :

$$W(G) = W(G_1) + W(G_2) + \dots + W(G_p)$$

nous nous intéresserons à l'utilisation de nos résultats avec ce type de graphe.

- Nous allons continuer de travailler sur l'indice de Wiener vu sa jeunesse et parce qu'il traite les extrêmes, ce qui va influencer peut être sur la recherche au niveau des réseaux informatiques.
- Nous allons également chercher d'autres regroupements de d'autres familles des indices topologiques.
- Et puisque les chercheurs de la conjecture de Wiener se basent dans la majorité des cas sur la formule 3.3, pour déterminer les graphes des valeurs particulières, et puisque aussi la formule 3.5 est sa généralisation nous allons travailler sur la conjecture de Wiener pour généraliser leurs méthodes.

Qu'est-ce que la classification ?

Les méthodes de classification ont pour but d'identifier les classes auxquelles appartiennent des objets définis à partir de certains traits descriptifs. Elles s'appliquent à un grand nombre d'activités humaines et conviennent en particulier au problème de la prise de décision automatisée. Il s'agira, par exemple, de prédire les propriétés et activités de molécules avant même que celles-ci ne soient synthétisées, d'établir un diagnostic médical à partir de la description clinique d'un patient, de donner une réponse à la demande de prêt bancaire de la part d'un client sur la base de sa situation personnelle, de déclencher un processus d'alerte en fonction de signaux reçus par des capteurs.

Dans ce présent chapitre que la majorité de ses paragraphes sont tirés de [93] nous allons présenter brièvement et à travers des exemples, ce qu'est un problème de classification. Nous allons constater que les problèmes de classification sont de deux types, pour certains de ces problèmes nous pouvons décrire explicitement, de manière linguistique, le mécanisme de classification. Dans ce cas, un algorithme reproduisant ce processus peut être construit et le problème est résolu. Pour d'autres problèmes, il est malheureusement impossible de décrire précisément la classification ; il faut alors trouver des méthodes de résolution qui «apprennent par exemple», à partir d'ensemble d'objets déjà classés par un expert.

1 Quelques exemples introductifs

Pour introduire et toucher de près la notion de classification, voyons ensemble ces exemples simples de classification qu'on les trouve dans notre vie quotidienne, et à travers desquels on comprend la notion de classification.

Exemple .1. Quelques exemples de classification :

- Un exemple courant d'application de la classification est le tri automatique du courrier par un dispositif de lecture et d'interprétation du code postal ou de l'adresse manuscrite. Pour

Qu'est-ce que la classification ?

un dispositif d'interprétation du code postal, 10 classes sont possibles (les chiffres de 0 à 9) et les variables descriptives peuvent être les niveaux de gris des pixels, provenant d'une image numérisée du chiffre à identifier.

- Un établissement bancaire est fréquemment appelé à répondre à la demande de prêt d'un client, sur la base de quelques indicateurs décrivant sa capacité à rembourser. Dans ce cas, les individus à classer sont des personnes, et les variables descriptives sont, par exemple, le salaire, l'âge, la situation de famille, le nombre d'enfants..., Nous pouvons imaginer plusieurs classes suivant le type de risque que peut admettre l'établissement.
- Pour un système de sécurité, le dispositif doit repérer au bon moment une situation inquiétante parmi la masse de situations normales et déclencher l'alerte. Dans le cas d'un dispositif de surveillance d'un réacteur chimique, les individus sont les états du processus au cours du temps et les variables descriptives sont, par exemple, la température, le débit, le *PH*... il y a deux classes possibles : situation normales ou bien situation anormale.

Dans ce mémoire, nous cherchons à calculer les descripteurs topologiques des molécules (ou système moléculaire), après avoir les transformées en des graphes planaires. Ces descripteurs vont être utilisés comme des variables descriptives, dans une méthode de classification pour classer les molécules selon les activités biologiques qui peuvent les faire. Donc, les objets qu'on cherche à classer sont les systèmes moléculaires, les variables sont les indices topologiques et les classes sont les activités biologiques qu'un système moléculaire (ou molécule) peut le provoquer. Pour illustrer le processus de la classification des molécules, nous allons indiquer dans le paragraphe suivant, comment la classification s'insère dans un système de reconnaissance de forme.

2 Le système de reconnaissance de forme

Puisque les objets qu'on manipule et on cherche à les décrire, pour les classer, durant ce mémoire, peuvent être vus comme des formes (les molécules); nous expliquerons dans ce paragraphe le dispositif de reconnaissance automatique de formes, qui est généralement conçu comme une chaîne de modules de traitement [58]. Ainsi, un système de reconnaissance de formes comporte habituellement :

2.1 Module d'acquisition

Dans ce module, les capteurs mesurent des grandeurs caractéristiques de l'objet à classer. Cet ensemble de grandeurs constitue la première représentation de l'objet.

2.2 Module de prétraitement

Dans ce module on trouve des algorithmes de filtrage, par ce qu'il se peut être judicieux de modifier les grandeurs brutes issues des capteurs par un algorithme afin de tenir compte des connaissances qui peuvent être disponibles a priori sur le problème. Par exemple de filtres destinés à annuler les effets de taille ou de positionnement pour la classification envisagée.

2.3 Modules complémentaires

Après le module de prétraitement, autres modules de traitement peuvent être se trouvés, qui élaborent les représentations successives de l'objet ; ces différentes représentations ont généralement pour objectif de réduire la dimension de la représentation, c'est-à-dire de diminuer le nombre de descripteurs de l'objet, et d'élaborer des descripteurs de plus en plus pertinents pour la tâche de classification à accomplir.

2.4 Module de classification

Le module le plus important dans ce processus est ce module. L'algorithme de classification, ou bien module de classification, se considère la dernière représentation de l'objet et décide d'affecter celui-ci à une classe. Cet algorithme peut fournir soit une réponse binaire à valeurs discrètes (appartenance ou non à une classe), soit une réponse binaire à valeurs continues (l'image à 70% de chance de représenter le chiffre 5).

2.5 Exemple d'un système de reconnaissance de formes

Exemple .2. La Figure A.1 illustre un exemple d'une chaîne de classification comporte un seul module de prétraitement. On rappelle juste que le module d'acquisition de cette chaîne, dans notre thèse la phase de la transformation d'une molécule (ou un système moléculaire) en un graphe planaire qui lui correspond, puis après cette transformation on calcule les descripteurs (indices) topologiques que nous les utiliserons par la suite dans la phase de classification. Dans la figure on distingue aussi les trois modules et les représentations successives de l'objet. Naturellement, on peut imaginer un dispositif sans module de prétraitement ; dans ce cas l'algorithme de classification travaille directement sur les grandeurs relevées par les capteurs.

Remarque .3. Dans le processus d'un système de reconnaissance de formes La tâche de l'algorithme de classification est d'autant plus aisée que la représentation de l'objet est pertinente. Par exemple dans un problème de commande d'un processus chimique, on peut imaginer que la distinction entre les situations normales et les situations de danger est entièrement définie

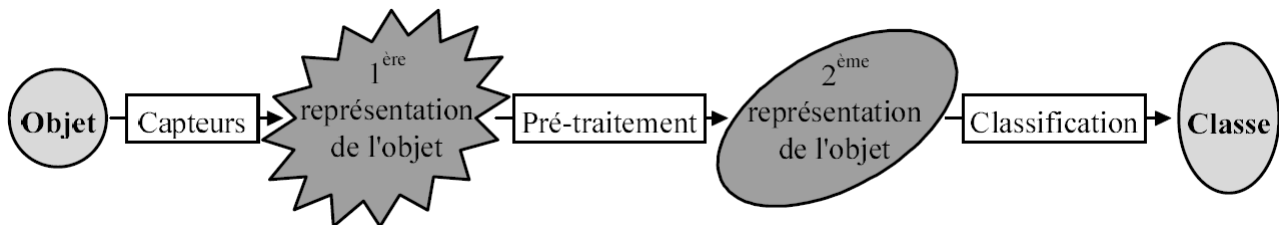


Figure A.1 – chaîne d'un dispositif de classification

par la valeur de pression. Si les modules d'acquisition ou de prétraitement ne fournissent pas cette valeur à l'algorithme de classification, celui-ci ne pourra pas faire de miracle et distinguer les différentes situations.

3 Le problème de classification mathématiquement

Les exemples précédents font apparaître la classification comme une tâche qui consiste à ranger des formes ou individus décrits par un ensemble de variables descriptives en un certain nombre de catégories ou classes définies à priori.

Traduit en termes mathématiques, un problème de classification comporte les ingrédients suivants :

- une population de N individus I^i , (i variant de 1 à N),
- P variables descriptives X_d^i , qui permettent de décrire les individus ; elles sont aussi appelées plus simplement descripteurs (variant de 1 à P),
- C classes C_k , dans lesquelles on cherche à ranger les individus (k variant de 1 à C),

Résoudre un problème de classification, c'est trouver une application de l'ensemble des objets à classer, décrits par les variables descriptives choisies, dans l'ensemble des classes. L'algorithme ou la procédure qui réalise cette application est appelé classifieur.

Les variables descriptives considérées ici sont celles qui sont fournies à l'algorithme de classification. Comme indiqué plus haut, elles peuvent être le résultat d'un prétraitement des variables initiales.

4 Quelques exemples distinctifs

Pour différencier les méthodes de classification existantes : **les méthodes paramétriques, non paramétrique, semi paramétriques, symboliques, non symboliques, probabilistes et non probabilistes** ; étudiant ensemble ces exemples de classification.

4.1 La classification exacte

4.1.1 Le rangement de pièces de monnaie

Nous trouvons un premier exemple de classification dans la vie de tous les jours : le rangement de pièces de monnaie. En effet, un commerçant doit, de temps à autre, rassembler les pièces identiques contenues dans sa caisse afin d'en faire des rouleaux qu'il remettra à la banque. Dans cet exemple, le fond de caisse du commerçant constitue la population concernée, chaque individu est une pièce de monnaie. Les classes sont au nombre de neuf :

Pièces de 5 centimes	Pièces de 10 centimes	Pièces de 20 centimes
Pièces de 50 centimes	Pièces de 1 Dirhams	Pièces de 2 Dirhams
Pièces de 5 Dirhams	Pièces de 10 Dirhams	pièces étrangères ou fausses

Les variables descriptives sont nombreuses, on trouve par exemple : Diamètre, épaisseur, poids, couleur(s), matériau (composition chimique), mot/chiffres/dessins en relief à la surface, bruit qui fait la pièce en tombant,... Ces descripteurs peuvent être considérés comme des grandeurs descriptives potentielles. Dans notre exemple, chacun d'entre eux est pertinent pour départager les pièces. Cependant, il n'est pas nécessaire de les utiliser tous. En réalité, les descripteurs dont peut se servir le commerçant sont la couleur et le diamètre (même s'il ne le mesure pas, mais l'évalue seulement). Dans ce cas, la relation F 1 qui relie les variables descriptives à la classe est de la forme :

$$\begin{aligned}
 F(\text{jaune, petit diamètre}) &= \text{classe des 5 centimes,} \\
 F(\text{jaune, moyen diamètre}) &= \text{classe des 10 centimes,} \\
 &\dots = \dots, \\
 F(\text{blanc, petit diamètre}) &= \text{classe des 50 centimes,} \\
 F(\text{blanc et jaune, grand diamètre}) &= \text{classe des 10 dirhams.}
 \end{aligned}$$

Notons toutefois qu'une personne aveugle n'utiliserait ni ces descripteurs ni cette fonction mais peut-être une fonction G 1 telle que :

$$G(\text{diamètre, épaisseur, surface de la tranche}) = \text{classe de la pièce.} \quad (1)$$

On voit donc que plusieurs règles de décision, toutes aussi pertinentes les unes que les autres, permettent de ranger cette population dans les classes désirées. Dans cet exemple, les fonctions peuvent être décrites explicitement (le commerçant ou l'aveugle peuvent expliquer comment

Qu'est-ce que la classification ?

ils procèdent) et elles conduisent au même résultat. Lorsqu'une telle tâche doit être effectuée de manière répétitive, on est tenté de la confier à un automate (c'est d'ailleurs le cas dans les caisses automatiques de parkings, distributeurs de titres de transport, etc.). En effet, dès que les variables descriptives et la fonction peuvent être exprimées si simplement, une telle classification «mécanique» peut facilement être réalisée par un automate réalisant une suite d'opérations logique (système expert). Celui-ci se fondera peut-être sur le gabarit, le poids ou la composition chimique des pièces, c'est-à-dire utilisera la fonction suivante :

$$H(\text{poids}, \text{diamètre}, \text{composition chimique}) = \text{classe de pièce}. \quad (2)$$

Malheureusement, les processus de classification ne sont pas toujours aussi simples et la règle de décision ne peut pas toujours être explicitée.

4.1.2 La notation des restaurants dans les guides touristiques

D'autres tâches de classification, qui sont, elles, fondées sur l'intuition, sont susceptibles d'être automatisées. La notation des restaurants dans les guides touristiques est, par exemple, un problème de classification plus complexe. Il s'agit bien de classer n'importe quel restaurant dans l'une des quatre classes : aucune étoile, une étoile, deux étoiles ou trois étoiles. En essayant soi-même d'évaluer tous les restaurants où l'on a déjà mangé (en prenant quatre niveaux : exceptionnel, satisfaisant, correct et à éviter) puis d'expliquer sa propre classification.

Remarque .4. on constate que :

- il n'est pas toujours facile de faire la liste des éléments que l'on prend en considération (les variables descriptives),
- il est quasiment impossible de formaliser la règle de décision que l'on adopte, c'est-à-dire de décrire comment s'élabore notre jugement. Dans un cas, le sourire de la serveuse aura suffi à compenser la tiédeur du steak et la table sera classée «correcte»; une valeur très positive de variable «service» aura prédominé sur la «piètre qualité du repas». Dans une autre circonstance, un délicieux foie gras fera oublier qu'on l'a attendu trois quarts d'heure en contemplant des murs lépreux ; la variable «qualité du repas» l'a emporté sur deux variables «service» et «cadre», etc.

Ainsi, la classification est souvent complexe dans les problèmes pour lesquels l'expert réagit en fonction de son intuition et ne peut pas toujours formaliser la fonction qu'il adopte. Pourtant, il peut être nécessaire de savoir reproduire la classification de l'expert. Par exemple, les chargés de clientèle d'une banque ne peuvent pas se contenter systématiquement d'une évaluation subjective et personnelle de solvabilité d'un client qui leur demande un prêt. Or, donner un

avis favorable ou défavorable à la demande du client revient à effectuer une classification des demandes en deux classes : celles que l'on accepte et celles que l'on refuse. L'image de marque de la banque et sa sécurité financière exigent que cette classification soit unifiée, dans toute la mesure du possible.

4.2 La classification probabiliste

4.2.1 La discrimination des femmes des hommes

Dans les exemples précédents, la classe des individus est bien définie ; mais ce n'est pas toujours le cas. Considérons une autre tâche qui consiste à discriminer les femmes des hommes à partir du seul facteur : taille. Pour simplifier, supposons que, l'on dispose des deux éléments suivants :

- Il y a autant de femmes que d'hommes dans la population considérée
- Après la croissance, les femmes adultes mesurent en moyenne 1,65 m avec un écart-type de 16 cm (moyenne égale à 1,75 m et écart-type égale à 15 cm pour les hommes). Ces données n'ont bien évidemment aucune valeur significatives ; juste à titre d'exemple.

La question dont nous chercherons sa réponse est : quelle est le sexe d'une personne mesurant 1,60 m ?

Comment répondre intelligemment à cette question ? Une première réflexion de bon sens conduit à dire que cette personne est une femme. Mais, tout le monde connaît des hommes de cette taille. La réponse est donc erronée. Une meilleure réponse consistera à dire, par exemple, que cet individu a une probabilité de 60% d'être une femme et la probabilité complémentaire d'être un homme (40%).

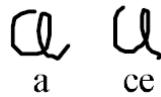
Nous n'avons plus à faire à une classification binaire (c'est une femme ou c'est un homme) mais à une classification probabiliste. De plus, face à un tel problème, une réponse probabiliste est une bonne solution ; en effet, la taille ne suffit pas à départager distinctement les deux classes, mais elle apporte une information interprétée en termes de probabilité.

5 Résolution des problèmes de classification

Lorsque l'expert ne peut pas expliciter son processus de classification, il faut se tourner vers des systèmes de classification qui «apprennent par exemple ». A partir d'un lot d'individus déjà classés par l'expert, le système peut apprendre à classer comme l'expert. Après apprentissage, le système est capable de classer de nouveaux individus.

5.1 Le principe de résolution

Prenons l'exemple de la lecture qui est aussi un exercice de classification. En effet, elle consiste, pour un texte normal, à classer des signes en 26 classes que sont les lettres de l'alphabet. Si la classification sous-jacente à toute lecture ne pose pas beaucoup de problèmes lorsqu'il s'agit d'un document imprimé, on sait à quel point l'exercice peut devenir difficile avec certaines écritures manuscrites. Par exemple, les signes ci-dessous doivent-ils être lus «a »ou «ce »? Dans la pratique, le contexte permet d'élucider la plupart de ces difficultés de déchiffre d'une



écriture, c'est-à-dire de classification des signes qui la composent. Mais, lorsque le sens ne permet pas cette élucidation, il reste la possibilité de regarder comment sont écrits les autres «a » que l'on a reconnus de manière certaine.

Ce petit exemple illustre le principe de résolution des problèmes de classification à partir d'observations, que nous désignons, conformément à l'usage dans le domaine des réseaux de neurones, sous le terme d'exemples. Pendant la phase d'apprentissage, on apprend à reconnaître la lettre «a » dans quelques cas non ambigus, et, par la suite, on peut identifier ce signe dans d'autres situations.

5.2 La procédure de résolution par apprentissage

La résolution des problèmes de classification par apprentissage se déroule donc en plusieurs étapes :

5.2.1 Première étape

faire classer un échantillon d'individus par un expert ; cet échantillon est désigné, dans le domaine des réseaux de neurones, sous le nom de base d'apprentissage,

5.2.2 Deuxième étape

concevoir et mettre en uvre un algorithme (appelé classifieur) qui parvient à reproduire la classification de l'échantillon d'apprentissage,

5.2.3 Troisième étape

évaluer la qualité du classifieur en l'appliquant à un ensemble d'individus classés par expert, mais qui n'ont pas été utilisés au cours de la phase d'apprentissage (cet ensemble est la base de test),

5.2.4 Dernière étape

si le test est satisfaisant, appliquer la méthode de deuxième étape à l'ensemble de la population à classer.

C'est cette procédure qui est appliquée lorsqu'on confie la résolution d'un problème de classification à une machine. Elle porte alors le nom de classification supervisée car elle requiert l'intervention d'un « superviseur » ou expert. Notons dès maintenant que la deuxième étape consiste bien à reproduire la classification de la base d'apprentissage à l'aide d'un algorithme numérique, et non pas à expliquer de manière linguistique la règle de décision mise en œuvre.

5.3 Les précautions à prendre en compte lors de l'apprentissage

En pratique, la procédure décrite précédemment peut se révéler à mettre en œuvre et conduire à un résultat inexploitable. Prenons l'exemple fictif d'une population d'individus $\{I\}$ que l'on voudrait ranger dans trois classes A , B , et C . Imaginons la situation suivante : un expert est sollicité pour effectuer cette classification et, après avoir classé les 15 premiers individus, il démissionne sans transmettre son savoir-faire. Afin de poursuivre ce travail, l'entreprise demande à un de ses employés de remplacer l'expert. Le collègue, qui découvre cette étude, commence par recueillir 3 grandeurs (X , Y et Z) caractéristiques des 15 individus. Il lui faut maintenant retrouver la fonction de classification de l'expert. Ces 15 individus constituent donc la base d'apprentissage pour le collègue. Sa première tâche va consister à essayer de reproduire la classification de cet échantillon (deuxième étape de la procédure). En l'occurrence, il va tenter de trouver quelle réflexion l'expert a bien pu mener pour conclure que le premier individu I^1 décrit par $(1; 2; 4)$ devait être rangé en classe A . Comme dans l'exemple des pièces, on se rend rapidement compte que plusieurs règles de décision peuvent réaliser la classification de l'échantillon (car il existe une infinité de fonctions capables de reproduire exactement la classification de n'importe quel nombre fini d'individus). La première à la quelle on peut penser est la fonction

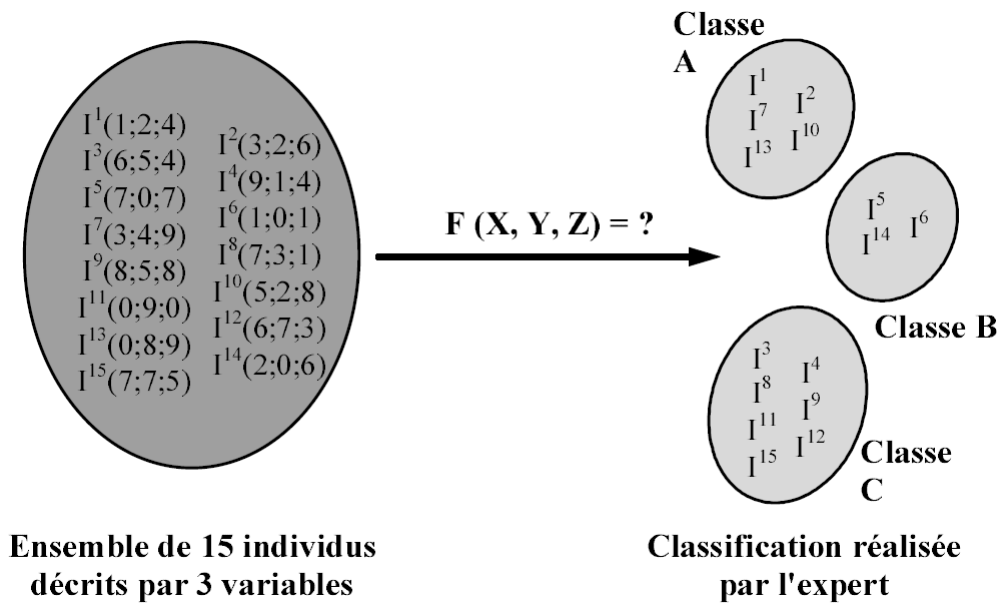


Figure A.2 – Un problème de classification

F_1 3 suivante :

$$F_1(X, Y, Z) = \begin{cases} A, & \text{Si } X = 1, Y = 2 \text{ et } Z = 4; \\ A, & \text{Si } X = 3, Y = 2 \text{ et } Z = 6; \\ \dots, & \dots; \\ C, & \text{Si } X = 7, Y = 7 \text{ et } Z = 5. \end{cases} \quad (3)$$

F_1 3 énonce, individu par individu, la classification de l'expert qu'elle reproduit donc parfaitement. Mais, si l'individu suivant est décrit par (2; 0; 0), dans quelle classe se range-t-il ? Cette première fonction envisagée ne permet pas de répondre. On peut dire qu'elle est top spécialisée.

Considérons une autre fonction, F_2 4, définie par :

$$F_2(X, Y, Z) = \begin{cases} A, & \text{Si } X + Y < Z; \\ B, & \text{Si } X + Y = Z; \\ C, & \text{Si } X + Y > Z. \end{cases} \quad (4)$$

F_2 4 reproduit parfaitement la classification de l'expert à une exception près, l'individu I^{14} décrit par (2; 0; 6). F_2 4 le range en classe A (2 + 0 < 6) alors que l'expert l'a mis en classe B. Cependant, cette deuxième fonction, plus générale, présente l'intérêt considérable de permettre de classer tout nouvel individu : c'est un modèle prédictif alors que la méthode précédente constituait seulement un modèle descriptif.

Alors que son collègue vient juste de trouver cette fonction F_2 4 qui semble satisfaisante, l'expert revient et lui apprend qu'en fait la fonction de classification ne portait que sur Y , qu'il ne fallait tenir compte ni de X ni de Z figurant dans le dossier pour d'utilisations. Sa règle est en effet :

$$F(X, Y, Z) = \begin{cases} A, & \text{Si } Y \text{ est pair;} \\ B, & \text{Si } Y = 0; \\ C, & \text{Si } Y \text{ est impair.} \end{cases} \quad (5)$$

Cet exemple illustre, de façon caricaturale, les principales difficultés que l'on rencontre dans la résolution par apprentissage.

Problème .1. Donc, on peut dire que les problèmes de la classification supervisée sont :

- Choix des variables descriptives : dans l'exemple, si la seule variable Y avait figuré, la règle de classement fondée sur la parité aurait vraisemblablement sauté aux yeux !
- Optimisation de la fonction : il faut toujours trouver un compromis entre une fonction très performante sur les individus de la base d'apprentissage et une fonction peut-être moins performante sur l'échantillon, mais qui présente de meilleures capacités de « généralisation ».
- Taille de l'échantillon : si la classe B avait comporté 50 individus, on aurait certainement vu que leur point commun était d'avoir une valeur de Y nulle. Autrement dit, la base d'apprentissage doit être suffisamment grande et représentative.

La suivante section résumera tous ce que nous venons de le discuter dans ce présent chapitre. Elle donnera la notion pratique de la classification, ses types, et quelques domaines d'application.

6 Qu'est-ce que la classification ?

6.1 Système expert

Comme nous l'avons vu durant ce brève chapitre, les approches utilisées pour résoudre le problème de classification sont nombreuses, dont la première est l'approche « systèmes experts », dans laquelle la connaissance d'un expert est décrite sous forme de règles. Cet ensemble de règles forme un système expert qui est utilisé pour classer de nouveaux cas. Cette approche, qui était largement utilisée dans les années 80, dépend fortement de la capacité à extraire et à formaliser les connaissances de l'expert.

6.2 Système d'apprentissage

Vu la difficulté de mathématiser plusieurs descripteurs qu'on peut les considérer (personnelles à l'expert), les études se dévient vers l'approche de «système d'apprentissage » pour laquelle la procédure de classification sera extraite automatiquement à partir d'un ensemble d'exemples. Un exemple consiste en la description d'un cas avec la classification correspondante. Un système d'apprentissage doit alors, à partir de cet ensemble d'exemples, extraire une procédure de classification qui, au vu des variables descriptives d'exemple, devra être capable de classer des nouveaux cas. Il s'agit donc d'induire une procédure de classification générale à partir d'exemples. Le problème est donc un problème inductif, il s'agit en effet d'extraire une règle générale à partir de données observées. La procédure générée devra classer correctement les exemples de l'échantillon mais surtout avoir un bon pouvoir prédictif pour classer correctement de nouvelles descriptions.

Les méthodes utilisées par les systèmes d'apprentissage sont très nombreuses et sont issues de domaines scientifiques variés.

6.2.1 Les méthodes statistiques

Les méthodes statistiques supposent que les descriptions des objets d'une même classe se répartissent en respectant une structure spécifique à la classe. On fait des hypothèses sur les distributions des descriptions à l'intérieur des classes et les procédures de classification seront construites à l'aide d'hypothèses probabilistes. La variété des méthodes viendra de la diversité des hypothèses possibles. Ces méthodes sont appelées semi-paramétriques. Des méthodes non paramétriques (sans hypothèse a priori sur les distributions) ont été également proposées en statistiques.

6.2.2 Les méthodes non statistiques

Les méthodes issues de l'intelligence artificielle sont des méthodes non paramétriques. On distingue les méthodes symboliques (la procédure de classification produite peut être écrite sous forme de règles) des méthodes non symboliques ou adaptatives (la procédure de classification produite est de type «boîte noire»). Parmi les méthodes symboliques, les plus utilisées sont basées sur les arbres de décision. Pour les méthodes adaptatives, on distingue deux grandes classes : les réseaux de neurones et les algorithmes génétiques.

6.2.3 Les applications de l'apprentissage

L'apprentissage automatique, dans une définition très générale, consiste en l'élaboration des programmes qui s'améliorent avec l'expérience. Les applications sont nombreuses et concernent des domaines très variés. On peut citer, par exemple, la reconnaissance de formes avec, en particulier, la reconnaissance de la parole et du texte écrit, le contrôle de processus et le diagnostic de pannes, les programmes de jeu.

Les méthodes d'apprentissage à partir d'exemples sont très utilisées dans la recherche d'informations dans de grands ensembles de données. En effet, l'évolution de l'informatique permet de nos jours de manipuler des ensembles de données de très grande taille (datawarehouse ou entrepôt de données). Par exemple, les chaînes de magasin peuvent mémoriser de grandes quantités de données concernant les consommateurs et ce qu'ils achètent. Le développement des technologies Internet et Intranet font que de nombreuses données issues de sources diverses et dans des formats variés deviennent accessibles.

Le processus de recherche d'informations dans de grands ensembles de données (KDD : Knowledge Discovery in Databases) comporte différentes étapes :

1. la sélection des données (extraction des informations de l'entrepôt) ;
2. la préparation des données (suppression des doublons, élimination des données aberrantes, ...);
3. le codage des données (normalisation des données, choix de codage,...);
4. la phase d'extraction proprement dite appelée fouille de données (Data mining) ;
5. la sortie des résultats.

La phase d'extraction d'information utilise les outils usuels d'interrogation tels que les requêtes SQL standard et les requêtes multi-dimensionnelles, mais aussi, pour l'extraction d'informations cachées, les algorithmes d'apprentissage à partir d'exemples. Les algorithmes les plus utilisés sont : les k -plus proches voisins, les arbres de décision, les systèmes de règles (programmation logique inductive), les réseaux de neurones et les algorithmes génétiques. Citons, parmi d'autres, quelques applications et leur domaine :

- Chimie : les méthodes de QSAR/QSPR, les méthodes de criblage virtuel,
- Analyse financière : prévision d'évolution de marchés,
- Banque : attribution de prêts,
- Médecine : aide au diagnostic, découvert des nouveaux médicaments.
- Télécom : détection de fraudes.

Qu'est-ce que la classification ?

Références bibliographiques

- [1] *International Conference on Discrete Mathematics and Computer Science (DIMACS'11)*, 2011. [2](#), [45](#), [98](#)
- [2] *International Workshop On Information Technologies and Communication (WO-TIC2011)*, 2011. [2](#), [58](#), [63](#), [64](#), [96](#)
- [3] *The 3rd International Conference on Multimedia Computing and Systems (ICMCS'12)*, 2012. [2](#), [49](#), [56](#), [57](#), [58](#), [59](#), [63](#), [64](#), [83](#), [84](#), [86](#), [92](#), [93](#), [95](#), [96](#)
- [4] *International Symposium On Signal, Image, Video and Communications (ISIVC2012)*, 2012. [2](#), [49](#), [53](#), [55](#), [56](#), [58](#), [83](#), [84](#), [85](#), [92](#), [93](#), [95](#), [96](#)
- [5] *Journées Doctorales en Système d'Information, Réseaux et Télécommunication (JD-SIRT2012)*, 2012. [2](#), [49](#), [53](#), [54](#), [55](#), [83](#), [84](#), [85](#), [92](#), [93](#), [95](#)
- [6] *Journées Doctorales en Système d'Informatique et Télécommunication (JDSIT'12)*, 2012. [2](#), [56](#), [57](#), [59](#), [61](#), [62](#), [86](#), [95](#), [96](#)
- [7] *Palestinian International Conference on Information and Communication Technology (PICICT2013)*, 2013. [2](#), [49](#), [56](#), [57](#), [58](#), [59](#), [61](#), [62](#), [83](#), [84](#), [86](#), [92](#), [93](#), [95](#), [96](#)
- [8] A.A.Dobrynin and A.A.Kochetova. Degree distance of a graph a degree analogue of the wiener index. *J. Chem. Inf. Comput. Sci.*, 34(5) :1082–1086, 1994. [87](#), [89](#), [90](#)
- [9] A.Cayley. On the mathematical theory of isomers. *Philos. Mag.*, 10(47) :444–446, 1874. [20](#), [114](#)
- [10] A.Cayley. On the mathematical theory of isomers. *Philos. Mag.*, 10 :444–446, 1874. [113](#)
- [11] A.Cheng, D.Diller, S.Dixon, W.Egan, G.Lauri, and K.Merz. Computation of the physio-chemical proprieties and data mining of large molecular collections. *J. Comput. Chem.*, 23(1) :172–183, 2002. [37](#)

Références bibliographiques

- [12] A.Crum-Brown and T.Frazer. On the connection between chemical constitution and physiological action. *Transactions of the Royal Society of Edinburgh*, 25 :151–203, 1868-69. [25](#)
- [13] A.Dobrynin. Branchings in trees and the calculation of the wiener index of a tree. *MATCH Commun. Math. Comput. Chem.*, 41 :119–134, 2000. [67](#)
- [14] A.Dobrynin and I.Gutman. On a graph invariant related to the sum of all distances in a graph. *Publ. Inst. Math. (Beograd)*, 56 :18–22, 1994. [65](#), [66](#)
- [15] A.Dobrynin, R.Entringer, and I.Gutman. Wiener index of trees : theory and application. *Acta Appl. Math.*, 66 :211–249, 2001. [71](#), [72](#), [75](#), [76](#)
- [16] A.GOULON-SIGWALT-ABRAM. *Une nouvelle méthode d'apprentissage de données structurées : applications à l'aide à la découverte de médicaments [thèse en ligne]*. PhD thesis, Paris : Université Pierre et Marie Curie (Paris 6), 2008. [1](#), [22](#), [23](#), [24](#), [25](#), [30](#), [31](#), [32](#)
- [17] A.Graovac, I.Gutman, and N.Trinajstè. *Topological Approach to the Chemistry of Conjugated Molecules*. Springer-Verlag, Berlin, 1977. [19](#)
- [18] A.Graovac, I.Gutman, and N.Trinajstè. *Topological Approach to the Chemistry of Conjugated Molecules*. 1977. [46](#)
- [19] A.Heydari and I.Gutman. On the terminal wiener index of thorn graphs. *Kragujevac J. Sci.*, 32 :57–64, 2010. [47](#)
- [20] A.Heydari and I.Gutman. On the terminal wiener index of thorn graphs. *Kragujevac J. Sci.*, 32 :57–64, 2010. [99](#), [100](#)
- [21] A.Höskuldson. Pls regression methods. *Journal of Chemometrics*, 2 :211–228, 1988. [28](#)
- [22] A.Jain, J.Mao, and K.Mohiuddin. Artificial neural networks : a tutorial. *Computer*, 29(3) :31–44, 1996. [30](#)
- [23] A.Leo and al. Calculation of hydrophobic constant ($\log p$) from π and f constants. *Journal of Medicinal Chemistry*, 18 :865, 1975. [31](#)
- [24] A.T.Balaban. *Chemical Application of Graph Theory*. Academic Press, London, 1976. [19](#)
- [25] A.T.Balaban. Highly discriminating distance-based topological index. *Chemical Physics Letters*, 89(5) :399–404, 1982. [23](#)
- [26] A.T.Balaban, I.Motoc, D.Bonchev, and O.Mekenyan. Topological indices for structure-activity correlations. *Top. Curr. Chem*, 114 :21–55, 1993. [46](#)
- [27] A.T.Balaban, J.W.Kennedy, and L.V.Quintas. The number of alkanes having n carbons and a longest chain of length d. *Journal of Chemical Education*, 65 :304–313, 1988. [19](#)

- [28] A.Z.Dudek, T.Arodz, and J.Gálvez. Computational methods in developing quantitative structure-activity relationships (qsar) : A review. *Combinatorial Chemistry & High Throughput Screening*, 9 :213–228, 2006. [23](#), [26](#)
- [29] B.B.West. *Introduction to graph theory*. Pearson Education (Singapore), 2002. [7](#), [10](#)
- [30] B.Furtula, A.Graovac, and D.Vukičević. Augmented zagreb index. *J. Math. Chem.*, 48 :370–380, 2010. [126](#)
- [31] B.Horvat, T.Pisanski, and M.Randić. Terminal polynomials and star-like graphs. *MATCH Commun. Math. Comput. Chem.*, 60 :493–512, 2008. [98](#)
- [32] B.Lin, H.Cheng, J.Yang, and F.Xia. The wiener index of the corona two graphs $c_m \circ c_n$. *South Asian Journal of Mathematicas*, 2(2) :122–125, 2012. [59](#)
- [33] B.Liu and Z.You. A survey on comparing zagreb indices. *MATCH Commun. Math. Comput. Chem.*, 65(3) :581–593, 2011. [126](#)
- [34] C.Berge. *Graphes et hypergraphes*. 1970. [7](#), [10](#)
- [35] C.Cortes and V.Vapnik. Support-vector networks. *Mach. Learn.*, 20 :273–297, 1995. [30](#)
- [36] C.Hansch, A.Leo, and D.Hoekmann. *Exploring QSAR : hydrophobic, electronic and steric constants*. Washington, DC : American Chemical Society, 1995. [2](#), [23](#)
- [37] C.Senese, J.Duca, D.Pan, A.Hopfinger, and Y.Tseng. 4d-fingerprints, universal qsar and qspr descriptors. *J. Chem. Inf. Model.*, pages 1526–1539, 2004. [30](#)
- [38] D.Bonchev. *Information Theoretic Indices for Characterization of Chemical Structure*. Research Studies Press, 1983. [19](#)
- [39] D.Bonchev. *Information Theoretic Indices for Characterization of Chemical Structure*. Chichester, 1983. [46](#)
- [40] D.Bonchev and D.H.Rouvray. *Chemical Graph Theory Introduction and Fundamentals*. Gordon & Breach, 1991. [19](#)
- [41] D.Bonchev and D.H.Rouvray. *Chemical Graph Theory Introduction and Fundamentals*. Gordon & Breach, 1991. [20](#)
- [42] D.Bonchev and D.H.Rouvray. *Chemical Graph Theory Introduction and Fundamentals*. Gordon Breach, New York, 1991. [46](#)
- [43] D.Chakrabarti. *Tools for Large Graph Mining*. PhD thesis, Center for Automated Learning and Discovery, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, 2005. [46](#)
- [44] D.H.Rouvray. The value of topological indices in chemistry. *Commun. Math. Comput. Chem. (MATCH)*, 1 :125–134, 1975. [42](#)

Références bibliographiques

- [45] D.H.Rouvray. Mathematics and computational concepts in chemistry. *Horwood, Chichester*, pages 295–306, 1986. [46](#)
- [46] D.H.Rouvray. Predicting chemistry from topology. *journal of scientific American*, 255 :36–43, 1986. [19](#)
- [47] D.H.Rouvray. The prediction of biological activity using molecular connectivity indices. *Acta Pharm. Jugosl.*, 36 :239–251, 1986. [47](#)
- [48] D.H.Rouvray. The challenge of characterizing branching in molecular species. *Discr. Appl. Math.*, 19 :317–338, 1988. [46](#)
- [49] D.H.Rouvray and B.C.Crafford. The dependence of physicochemical properties on topological factors. *South Afr. J. Sci.*, 72 :47–51, 1976. [42](#)
- [50] D.H.Rouvray and R.B.King. *Chemical Applications of Topology and Graph Theory*, pages 159–177. 1983. [46](#)
- [51] D.J.Klein and I.Gutman. Wiener-number-related sequences. *J. Chem. Inf. Comput. Sci.*, 39 :534–536, 1999. [82](#)
- [52] D.J.Klein, I.Lukovits, and I.Gutman. On the definition of the hyper-wiener index for cycle-containing structures. *J. Chem. Inf. Comput. Sci.*, 35 :50–52, 1990. [82](#)
- [53] D.J.Klein, Z.Mihalić, D.Plavšić, and N.Trinajstić. Molecular topological index : A relation with the wiener index. *J. Chem. InJ Comput. Sci.*, 32 :304–305, 1992. [87](#), [88](#)
- [54] D.Kitchen, H.Decornez, J.Furr, and J.Bajorath. Docking and scoring in virtual screening for drug discovery : methods and applications. *Nat. Rev. Drug Discov*, 3 :935–949, 2004. [35](#)
- [55] D.M.Cvetković and I.Gutman. Note on branching. *Croat. Chem. Acta*, 49 :115–121, 1977. [127](#)
- [56] D.Plavšić, S.Nikolić, N.Trinajstić, and Z.Mihalić. On the harary index for the characterization of chemical graphs. *J. Math. Chem.*, 12 :235–250, 1993. [79](#)
- [57] D.Plavšić, S.Nikolić, N.Trinajstić, and D.J.Klein. Relation between the wiener index and the schultz index for several classes of chemical graphs. *Croat. Chem. Acta.*, 66 :345–353, 1993. [87](#)
- [58] D.Price. *Classification probabiliste par réseaux de neurones : application à la reconnaissance de l'écriture manuscrite*. PhD thesis, l'Université pierre et Marie Curie-Paris VI, 1996. [144](#)
- [59] D.T.Stanton. Evaluation and use of bcut descriptors in qsar and qspr studies. *J. Chem. Inf. Comput. Sci.*, 39(1) :11–20, 1999. [24](#)

- [60] D.Vukičević, A.Miličević, S.Nikolić, J.Sedlar, and N.Trinajstić. Paths and walks in acyclic structures : plerographs versus kenographs. *ARKIVOC*, pages 33–44, 2005. [113](#), [114](#), [116](#), [117](#)
- [61] D.Vukičević and N.Trinajstić. Wiener indices of benzenoid graphs. *Bulletin of the Chemists and Technologists of Macedonia*, 23(2) :113–129, 2004. [47](#)
- [62] E.A.Smolenskii, E.V.Shuvalova, L.K.Maslova, I.V.Chuvaeva, and M.S.Molchanova. Reduced matrix of topological distance with a minimum number of independent parameters distance vectors and molecular codes. *J. Math. Chem.*, 45 :1004–1020, 2009. [98](#)
- [63] E.Bolton, Y.Wang, P.A.Thiessen, and S.H.Bryant. Chapter 12 pubchem : Integrated platform of small molecules and biological activities. *Annual Reports in Computational Chemistry*, 4 :217–241, 2008. [22](#)
- [64] E.Estrada. Characterization of the folding degree of proteins. *Bioinformatics*, 18(5) :697–704, 2002. [127](#)
- [65] E.Estrada. Atombond connectivity and the energetic of branched alkanes. *Chem. Phys. Lett*, 463 :422–425, 2008. [126](#)
- [66] E.Estrada. *The Structure of Complex Networks-Theory and Applications*. Oxford Univ. Press, 2012. [127](#)
- [67] E.Heilbronner. A simple equivalent bond orbital model for the rationalization of the c_{2s} -photoelectron spectra of the higher n -alkanes, in particular of polyethylene. *Helv. Chim. Acta*, 60 :2248–2257, 1977. [114](#)
- [68] E.R.Canfield, R.Robinson, and D.H.Rouvray. Determination of the wiener molecular branching index for the general tree. *J. Comput. Chem.*, 6 :598–609, 1985. [70](#)
- [69] E.Russo. Chemistry plans a structural overhaul. *Nature Jobs*, 419 :4–7, 2002. [21](#)
- [70] F.Bonachera. *Les triplets pharmacophoriques flous : développement et applications [thèse en ligne]*. PhD thesis, Lille : Université Lille1 sciences et technologies, 2011. [22](#), [23](#), [24](#), [25](#), [26](#), [35](#), [37](#)
- [71] F.Diot, E.Fromont, B.Jeudy, E.Marilly, and O.Martinot. Graph mining for object tracking in videos. *ECML PKDD 2012, Part I, LNCS*, 7523 :394–409, 2012. [46](#)
- [72] F.R.Burden. Molecular identification number for substructure searches. *Journal of Chemical Information and Computer Sciences*, 29 :225–227, 1989. [24](#)
- [73] F.Rosenblatt. The perceptron : A probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, 65(6) :386–408, 1959. [30](#)
- [74] F.Stahura and J.Bajorath. New methodologies for lignd-based virtual screening. *Curr. Pharm. Des*, 11(9) :1189–1202, 2005. [35](#)

Références bibliographiques

- [75] G.Bieri, J.D.Dill, E.Heilbronner, and A.Schmelzer. Application of the equivalent bond orbital model to the c_{2s} -ionization energies of saturated hydrocarbons. *Helv. Chim. Acta*, 60 :2234–2247, 1977. [114](#)
- [76] G.Klopman and al. Computer automated logp calculations based on an extended group approach. *Journal of Chemical Information and Computer Sciences*, 34(4) :752–781, 1994. [31](#)
- [77] G.Al Hagri, M.El Marraki, and M.Essalih. The wiener index of some particular graph. *International Journal of Computer Science and Network Security (IJCSNS)*, 11(5) :64–69, 2011. [2](#), [45](#), [98](#)
- [78] G.Al Hagri, M.El Marraki, and M.Essalih. The degree distance of certain particular graphs. *Applied Mathematical Sciences (AMS)*, 6(18) :857–867, 2012. [97](#)
- [79] H.Eckert and J.Bajorath. Molecular similarity analysis in virtual screening : foundations, limitations and novel approaches. *Drug Discov*, 12(5-6) :225–233, 2007. [35](#)
- [80] H.H.Günthard and H. Primas. Zusammenhang von graphentheorie und mo-theorie von molekeln mit systemen konjugierter bindungen. *Helv. Chim. Acta*, 39 :1645–1653, 1956. [114](#)
- [81] H.Hosoya. Topological index. a newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Japan*, 44 :2332–2339, 1971. [42](#)
- [82] H.Martens and T.Næs. *Multivariate calibration*. Chichester : Wiley, 1989. [28](#)
- [83] H.Narumi and H.Kita. Equivalent bond index. *Commun. Math. Chem. (MATCH)*, 30 :225–242, 1994. [114](#)
- [84] H.P.Schultz. Topological organic chemistry. 1. graph theory and topological indices of alkanes. *Journal of Chemical Information and Computer Sciences*, 29(3) :227–228, 1989. [23](#)
- [85] H.P.Schultz. Topological organic chemistry. 1. graph theory and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.*, 29 :227–228, 1989. [86](#), [87](#)
- [86] H.P.Schultz, E.B.Schultz, and T.P.Schultz. Topological organic chemistry. 2. graph theory, matrix determinants and eigenvalues, and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.*, 30 :27–29, 1990. [87](#)
- [87] H.P.Schultz, E.B.Schultz, and T.P.Schultz. Topological organic chemistry. 4. graph theory, matrix permanents, and topological indices of alkanes. *J. Chem. InJ Comput. Sci*, 32 :69–72, 1992. [87](#)

- [88] H.P.Schultz, E.B.Schultz, and T.P.Schultz. Topological organic chemistry. 7. graph theory and molecular topological indices of unsaturated and aromatic hydrocarbons. *J. Chem. Inf. Comput. Sci.*, 33 :863–867, 1993. [87](#), [88](#)
- [89] H.P.Schultz and T.P.Schultz. Topological organic chemistry. 3. graph theory, binary and decimal adjacency matrices, and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.*, 31 :144–147, 1991. [87](#)
- [90] H.P.Schultz and T.P.Schultz. Topological organic chemistry. 5. graph theory, matrix hafnians and pfaffians, and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.*, 32 :364–366, 1992. [87](#)
- [91] H.P.Schultz and T.P.Schultz. Topological organic chemistry. 6. theory and topological indices of cycloalkanes. *J. Chem. Inf. Comput. Sci.*, 33 :240–244, 1993. [87](#)
- [92] H.Stoppiglia. *Méthodes statistiques de sélection de modèles neuronaux; applications financières et bancaires [thèse en ligne]*. PhD thesis, Paris : Université Pierre et Marie Curie, 1997. [29](#)
- [93] H.Stoppiglia. *Méthodes statistiques de sélection de modèles neuronaux; applications financières et bancaires [thèse en ligne]*. PhD thesis, ESPCI ParisTECH, 1997. [143](#)
- [94] H.Stoppiglia and al. Ranking a random feature for variable and feature selection. *Journal of Machine Learning Research*, 3 :1399–1414, 2003. [29](#)
- [95] H.Wang and G.Yu. All but 49 numbers are wiener indices of trees. *Acta Appl. Math.*, 92 :15–20, 2006. [59](#)
- [96] H.Wiener. Correlation of heats of isomerization, and differences in heats of vaporization of isomers, among the paraffin hydrocarbons. *J. Am. Chem. Soc.*, 69 :2636–2638, 1947. [42](#), [46](#)
- [97] H.Wiener. Influence of interatomic forces on paraffin properties. *J. Chem. Phys.*, 15 :766–766, 1947. [42](#), [46](#)
- [98] H.Wiener. Structural determination of paraffin boiling points. *Journal of Chemical Information and Computer Sciences*, 69 :17–20, 1947. [23](#), [42](#), [43](#), [46](#), [65](#)
- [99] H.Wiener. Relation of the physical properties of the isomeric alkanes to molecular structure. surface tension, specific dispersion, and critical solution temperature in aniline. *J. Chem. Phys.*, 52 :1082–1089, 1948. [42](#), [46](#)
- [100] H.Wiener. Vapor pressure-temperature relationships among the branched paraffin hydrocarbons. *J. Chem. Phys.*, 52 :425–430, 1948. [42](#), [46](#)
- [101] H.Wold. Estimation of principal components and related models by iterative least squares, in multivariate analysis, krishnaiah, p.r., editor. *New York : Academic Press*, pages 391–420, 1966. [28](#)

Références bibliographiques

- [102] H.Y.Zhu, D.J.Klein, and I.Lukovits. Extensions of the wiener number. *J. Chem. Inf. Comput. Sci.*, 36 :420–428, 1996. [79](#)
- [103] I.Gutman. Selected properties of the schultz molecular topological index. *J. Chem. Inf. Comput. Sci.*, 34(5) :1087–1089, 1994. [87](#), [88](#), [89](#), [90](#), [91](#), [92](#)
- [104] I.Gutman. A property of the wiener number and its modifications. *Indian J. Chem.*, 36(A) :128–132, 1997. [48](#), [49](#), [77](#), [78](#), [79](#), [80](#)
- [105] I.Gutman. Distance in thorny graph. *Publ. Inst. Math(Beograd)*, 63 :31–36, 1998. [115](#)
- [106] I.Gutman. Distance of thorny graphs. *Publ. Inst. Math. (Beograd)*, 63 :31–36, 1998. [73](#), [74](#)
- [107] I.Gutman, A.A.Dobrynin, S.Klavžar, and L.Pavlović. Wiener-type invariants of trees and their relation. *Bull. Inst. Combin. Appl.*, 40 :23–30, 2002. [77](#), [78](#), [80](#), [141](#)
- [108] I.Gutman, B.Furtula, J.Tošović, M.Essalih, and M.El Marraki. On terminal wiener indices of kenograms and plerograms. *Iranian Journal of Mathematical Chemistry(IJMC)*, 4(1) :77–89, 2013. [3](#), [73](#), [74](#), [102](#), [103](#), [104](#), [105](#), [113](#), [114](#), [129](#), [133](#)
- [109] I.Gutman, B.Furtula, and M.Petrović. Terminal wiener index. *J Math Chem*, 46 :522531, 2009. [65](#), [66](#)
- [110] I.Gutman, B.Furtula, and M.Petrović. Terminal wiener index. *J Math Chem*, 46 :522–531, 2009. [98](#), [99](#), [100](#), [101](#), [102](#)
- [111] I.Gutman, B.Ručić, N.Trinajstić, and C.F.Wilcox. Graph theory and molecular orbitals. xii. acyclic polyenes. *J. Chem. Phys*, 62 :3399–3405, 1975. [88](#)
- [112] I.Gutman and D.Vidović. Two early branching indices and the relation between them. *Theor. Chem. Acc*, 108 :98–102, 2002. [127](#)
- [113] I.Gutman, D.Vidović, and L.Popović. Graph representation of organic molecules cayley’s plerograms vs. his kenograms. *J. Chem. Soc., Faraday Trans*, 94(7) :857–860, 1998. [113](#), [114](#), [115](#), [118](#), [119](#), [120](#), [121](#), [122](#), [129](#)
- [114] I.Gutman, D.Vidović, and L.Popović. On graph representation of organic molecules-cayley’s plerograms vs. his kenograms. *J. Chem. Soc. Faraday Trans.*, 94 :857–860, 1998. [77](#), [78](#)
- [115] I.Gutman and J.H.Potgieter. Performing your original search. *South Afr. J. Sci.*, 92 :47–48, 1996. [46](#)
- [116] I.Gutman and J.H.Potgieter. Wiener index and intermolecular forces. *J Serb. Chem. Soc.*, 62 :185–192, 1997. [46](#)

- [117] I.Gutman, M.Essalih, M.El Marraki, and B.Furtula. Why plerograms are not used in chemical graph theory? the case of terminal-wiener index. *Chemical Physics Letters*, 568 :195–197, 2013. [3](#), [19](#), [73](#), [74](#), [113](#), [114](#)
- [118] I.Gutman and O.E.Polansky. *Mathematical Concepts in Organic Chemistry*. Springer-Verlag, 1986. [19](#)
- [119] I.Gutman and O.E.Polansky. *Mathematical Concepts in Organic Chemistry*. Springer-Verlag, 1986. [20](#)
- [120] I.Gutman and O.E.Polansky. *Mathematical Concepts in Organic Chemistry*. Springer-Verlag, 1986. [114](#)
- [121] I.Gutman and Y.N.Yeh. The sum of all distances in bipartite graphs. *Mathematica Slovaca*, 45 :327–334, 1995. [76](#)
- [122] I.Gutman, Y.N.Yeh, and J.C.Chen. On the sum of all distances in graphs. *Tamkang J. Math.*, 25 :83–86, 1994. [47](#), [76](#)
- [123] I.Gutman, Y.N.Yeh, S.L.Lee, and Y.L.Luo. Some recent results in the theory of the wiener number. *Indian J. Chem*, 32(A) :651–661, 1993. [46](#)
- [124] I.Kuntz, J.Blaney, S.Oatley, R.Langridge, and T.Ferrin. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, 161 :269–288, 1982. [35](#)
- [125] I.Lukovits. Application to drugreceptor interactions. *J. Chem. Soc. Perkin II*, pages 1667–1671, 1988. [47](#)
- [126] I.Lukovits. Correlation between components of the wiener index and partition coefficients of hydrocarbons. *Intern. J. Quantum Chem. Quantum Biol. Symp.*, 19 :217–223, 1992. [47](#)
- [127] I.M.Kapetanovic. Computer-aided drug discovery and developement (caddd) : in silico-chemico-biological approach. *Chem Biol Interact*, 171(2) :76–165, 2008. [1](#)
- [128] I.T.Jolliffe. *Principal Component Analysis*. New-York, NY : Springer, 2ème édition, 2002. [28](#)
- [129] J.Bajorath. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Comput. Sci.*, 41(2) :233–245, 2001. [37](#)
- [130] J.Doyle and J.Graver. Mean distance in a graph. *Discrete Math*, 17 :147–154, 1977. [66](#), [67](#)
- [131] J.Gàlvez, R.Garcià-Domenech, M.T.Salabert-Salvador, and R.Soler. Charge indexes. new topological descriptors. *Journal of Chemical Information and Computer Sciences*, 34 :520–525, 1994. [24](#)

Références bibliographiques

- [132] J.H.Schuur, P.Selzer, and J.Gasteiger. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. *Journal of Chemical Information and Computer Sciences*, 36(2) :334–344, 1996. [25](#)
- [133] J.Labelle. *Théorie des graphes*. Bibliothèque nationale de Canada, 1981. [7](#), [10](#)
- [134] J.L.Durant, B.A.Leland, D.R.Henry, and J.G.Nourse. Reoptimization of mdl keys for use in drug discovery. *Journal of Chemical Information and Computer Sciences*, 42(6) :1273–1280, 2002. [24](#)
- [135] J.M.Barnard and G.M.Downs. Chemical fragment generation and clustering software. *Journal of Chemical Information and Computer Sciences*, 89-5 :141–142, 1997. [24](#)
- [136] J.Quinlan. Induction of decision trees. *Mach. Learn.*, 1 :81–106, 1986. [30](#)
- [137] J.R.Dias. *Molecular Orbital Calculations Using Chemical Graph Theory*. Springer-Verlag, 1993. [19](#)
- [138] J.Sadowski and J.Gasteiger. From atoms and bonds to three-dimensional atomic coordinates : automatic model builders. *Chemical Reviews*, 93 :2567–2581, 1993. [33](#)
- [139] J.V.Knop, W.R.Müller, K.Szymanski, and N.Trinajstić. On the determinant of the adjacency-plus-distance matrix as the topological index for characterizing alkanes. *J. Chem. InJ Comput. Sci.*, 31 :83–84, 1991. [87](#)
- [140] J.W.Jalowka and T.E.Daubert. Group contribution method to predict critical temperature and pressure of hydrocarbons. *Industrial and Engineering Chemistry Process Design and Development*, 25(1) :139–142, 1986. [32](#)
- [141] J.Xu. *Theory and Application of Graphs*. Kluwer Academic Publishers, 2003. [12](#)
- [142] K.Zaretskii. Reconstruction of a tree from the distances between its pendant vertices. *Uspekhi Math. Nauk.*, 20 :90, 1965. [98](#)
- [143] L.B.Kier and L.H.Hall. Derivation and significance of valence molecular connectivity. *Journal Pharm Sci.*, 70(6) :583–589, 1981. [24](#)
- [144] L.Gan, H.Hou, and B.Liu. Some results on atom-bond connectivity index of graphs. *MATCH Commun. Math. Comput. Chem.*, 66 :669–680, 2011. [126](#)
- [145] L.Stiel and G.Thodos. The normal boiling points and critical constants of saturated aliphatic hydrocarbons. *J. Am. Inst. Chem. Eng.*, 8 :527–534, 1962. [42](#)
- [146] L.Zhong. The harmonic index for graphs. *Appl. Math. Lett.*, 25 :561–566, 2012. [126](#)
- [147] M.A.Alipour and A.R.Ashrafi. A numerical method for computing the wiener index of one-heptagonal carbon nanocone. *Journal of Computational and Theoretical Nanoscience*, 6 :1–4, 2009. [47](#)

- [148] M.EL Marraki and A.Modabish. Wiener index of planar maps. *Journal of Theoretical and Applied Information Technology(JATIT)*, 18(1) :7–10, 2010. [47](#), [48](#), [49](#), [73](#)
- [149] M.EL Marraki and G.AL Hagri. Calculation of the wiener index for some particular trees. *Journal of Theoretical and Applied Information Technology(JATIT)*, 22(2) :77–83, 2010. [44](#), [45](#), [72](#)
- [150] M.Essalih, M.El Marraki, and G.Al hagri. Calculation of some topological indices. *Journal of Theoretical and Applied Information Technology (JATIT)*, 30(2) :122–127, 2011. [2](#), [44](#), [48](#), [49](#), [52](#), [54](#), [55](#), [56](#), [57](#), [83](#), [84](#), [85](#), [86](#), [92](#), [93](#), [94](#), [95](#)
- [151] M.Essalih, M.El Marraki, and G.Al hagri. Some topological indices of spider’s web planar graph. *Applied Mathematical Sciences(AMS)*, 6(63) :3145–3155, 2012. [2](#), [44](#), [49](#), [52](#), [54](#), [55](#), [56](#), [57](#), [58](#), [63](#), [64](#), [83](#), [84](#), [85](#), [86](#), [92](#), [93](#), [94](#), [95](#), [96](#)
- [152] M.Lepović and I.Gutman. A collective property of trees and chemical trees. *J. Chem. Inf. Comput. Sci.*, 38 :823–826, 1998. [76](#)
- [153] M.Randić. Characterization of molecular branching. *J. Am. Chem. Soc.*, 97(23) :6609–6615, 1975. [23](#)
- [154] M.Randić. On characterization of molecular branching. *J. Am. Chem. Soc.*, 97 :6609–6615, 1975. [126](#)
- [155] M.Randić. *Design of Molecules with Desired Properties. A Molecular Similarity Approach to Property Optimization, in Concepts and Applications of Molecular Similarity*, chapter 5, pages 77–145. John Wiley & Sons, 1990. [46](#)
- [156] M.Randić. Novel molecular descriptor for structure-property studies. *Chem. Phys. Lett.*, 211 :478–483, 1993. [81](#)
- [157] M.Randić. 2-d graphical representation of proteins based on virtual genetic code. sar-qsar. *Environ. Res.*, 15 :191–205, 2004. [98](#)
- [158] M.Randić, J.Zupan, and D.V.Topić. On representation of properties by starlike graphs. *J. Mol. Graph. Modell.*, 26 :290–305, 2007. [98](#)
- [159] M.Rigo. Théorie des graphes, 2009. [7](#), [10](#), [12](#)
- [160] M.Sana and G.Leroy. Graph theory, electronic structures and reaction mechanisms. *J. Mol. Struct. (THEOCHEM)*, 109 :251–269, 1984. [114](#)
- [161] M.V.Diudea. Layer matrices in molecular graphs. *J. Chem. Inf. Comput. Sci.*, 34 :1064–1071, 1994. [46](#)
- [162] M.V.Diudea and I.Gutman. Wiener-type topological indices. *CROATICA CHEMICA ACTA*, 71(1) :21–51, 1998. [47](#)

Références bibliographiques

- [163] M.V.Diudea, I.Gutman, and L.Jantschi. *Molecular Topology*. Science publishers, 1999. [7](#), [10](#), [20](#), [22](#), [43](#), [77](#)
- [164] M.V.Diudea and O.Ivanciuc. *Molecular Topology*. Complex (in Romanian), 1995. [19](#)
- [165] M.V.Diudea and O.Ivanciuc. *Molecular Topology*. Complex, Cluj, Romanian, 1995. [46](#)
- [166] M.V.Diudea, O.Ivanciuc, S.Nikolić, and N.Trinajstić. Matrices of reciprocal distance, polynomials and derived numbers. *MATCH - Commun. Math. Comput. Chem.*, 35 :41–64, 1997. [78](#), [79](#)
- [167] N.L.Biggs, E.K.Lloyd, and R.J.Wilson. *Graph theory 1736 – 1936*. Oxford University Press, 1976. [20](#)
- [168] N.S.SCHMUCK. *The Wiener index of a graph*. PhD thesis, Graz University of Technology, 2010. [44](#), [47](#), [48](#), [49](#), [65](#), [66](#), [67](#), [68](#), [69](#), [70](#), [71](#), [72](#), [73](#), [74](#), [75](#), [76](#)
- [169] N.Trinajstić. *Chemical Graph Theory*. CRC, Boca Raton, Fl., 1992. [19](#)
- [170] N.Trinajstić. *Chemical Graph Theory*. CRC, Boca Raton, Fl, 1992. [20](#)
- [171] O.E.Polansky and D.Bonchev. The wiener number of graphs. i. general theory and changes due to some graph operations. *MATCH Commun. Math. Comput. Chem.*, 21 :133–186, 1986. [72](#)
- [172] P.Dankelmann, I.Gutman, S.Mukwembi, and H.C.Swart. On the degree distance of a graph. *Discrete Applied Mathematics*, 157 :2773–2777, 2009. [91](#)
- [173] P.G.Seybold, M.May, and U.A.Bagal. Molecular structure-property relationships. *Journal of Chemical Education*, 64 :575–582, 1987. [19](#)
- [174] P.J.Hansen and P.C.Jurs. Chemical applications of graph theory. *J. Chem. Educ.*, 65 :574–580, 1988. [46](#)
- [175] P.J.Hansen and P.C.Jurs. Chemical applications of graph theory. part i. fundamentals and topological indices. *Journal of Chemical Education*, 65 :574–580, 1988. [19](#)
- [176] P.Lyne. Structure-based virtual screening : an overview. *Drug Discov. Today*, 7(20) :1047–1055, 2002. [35](#)
- [177] R.D.Cramer, D.E.Patterson, and J.D.Bunce. Comparative molecular field analysis (comfa). 1. effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, 110(18) :5959–5967, 1988. [25](#)
- [178] R.D.Cramer, D.E.Patterson, and J.D.Bunce. Comparative molecular field analysis (comfa). 1. effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, 110(18) :5959, 1988. [33](#)

- [179] R.Diestel. *Graph theory*. Springer, 2000. [12](#), [13](#)
- [180] R.Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugen*, 7 :179–188, 1936. [30](#)
- [181] R.Frucht and F.Harary. On the corona two graphs. *Aequationes Math*, 4 :322–325, 1970. [76](#)
- [182] R.Sheridan and S.Kearsley. Why do we need so many chemical similarity search methods? *DDT*, 7(17) :903–911, 2002. [37](#)
- [183] S.Gelfand, C.Ravishankar, and E.Delp. An iterative growing and pruning algorithm for classification tree design. *IEEE Tran. Pattern Anal. Mach. Intell*, pages 163–174, 1991. [30](#)
- [184] S.G.Wagner. A class of trees and its wiener index. *Acta Appl. Math.*, 91 :119–132, 2006. [76](#)
- [185] S.H.Zadeh, A.Hamzeh, and A.R.Ashrafi. Extremal properties of zagreb coindices and degree distance of graphs. *Miskolc Mathematical Notes*, 11(2) :129–137, 2010. [91](#)
- [186] S.Nikolić, N.Trinajstić, and Z.Mihalić. The wiener index : Development. and applications. *Croat. Chem. Acta*, 68 :105–129, 1995. [46](#)
- [187] S.P.Eu, B.Y.Yang, and Y.N.Yeh. Generalized wiener indices in hexagonal chains. *International Journal of Quantum Chemistry*, 106(2) :426–435, 2006. [47](#)
- [188] S.S.Tratch, M.I.Stankevich, and J.Zefirov. Combinatorial models and algorithms in chemistry. the expanded wiener number-a novel topological index. *J. Comput. Chem.*, 11 :899–908, 1990. [81](#)
- [189] S.Trohalaki, R.Pachter, K.Geiss, and J.Frazier. Halogenated aliphatic toxicity qsars employing metabolite descriptors. *J. Chem. Inf. Model.*, pages 1186–1192, 2004. [30](#)
- [190] S.Wold, M.Sjöström, and L.Eriksson. Pls-regression : a basic tool of chemometrics. *Chemom. Intell. Lab. Syst*, 58 :109–130, 2001. [30](#)
- [191] S.Word, C.Albano, D.I.W.J., K.Esbensen, S.Hellberg, E.Johansson, and H.Sjöström. Pattern recognition : finding and using regularities in multivariate data. *SIAM J. Sci. Stat. Comput.*, 5 :735–743, 1984. [30](#)
- [192] T.Cover and P.Hart. Nearsset neighbor pattern classification. *IEEE Trans. Inform. Theory*, pages 21–27, 1967. [30](#)
- [193] T.Hou, W.Zhang, K.Xia, X.Qiao, and X.Xu. Adme evaluation in drug discovery. 5 correlation of caco-2 permeation with simple molecular properties. *J. Chem. Inf. Model.*, pages 1585–1600, 2004. [30](#)

Références bibliographiques

- [194] T.W.Heritage and al. Eva : A novel theoretical descriptor for qsar studies. *Perspectives in Drug Discovery and Design*, 9-11(0) :381–398, 1998. [25](#)
- [195] V.Chepoi. The wiener index and the szeged index of benzenoid systems in linear time. *J. Chem. Inf. Comput. Sci.*, 37 :752–755, 1997. [47](#)
- [196] V.Gineityte. Secular polynomials for chemical graphs of alkanes in terms of atoms and bonds and their spectral properties. *Int. J. Quantum Chem.*, 60 :743–752, 1996. [114](#)
- [197] V.KBalakrishnan. *Graph theory*. Mcgraw-hill companies, 1995. [7](#), [10](#), [12](#)
- [198] V.Svetnik, T.Wang, C.Tong, A.Liaw, R.Sheridan, and Q.Song. Boosting : An ensemble learning tool for compound classification and qsar modeling. *J. Chem. Inf. Model*, pages 786–799, 2005. [30](#)
- [199] W.R.Müller, K.Szymanski, J.V.Knop, and N.Trinajstić. Molecular topological index. *J. Chem. Inf. Comput. Sci.*, 30 :160–163, 1990. [87](#)
- [200] Y.E.A.Ban, S.Bereg, and N.H.Mustafa. A conjecture on wiener indices in combinatorial chemistry. *Algorithmica*, 40(2) :99–118, 2004. [76](#)
- [201] Y.Huang, B.Liu, and L.Gan. Augmented zagreb index of connected graphs. *MATCH Commun. Math. Comput. Chem.*, 67(2) :483–494, 2012. [126](#)
- [202] Z.Mihalić and N.Trinajstić. A graph theoretical approach to structure-property relationships. *Journal of Chemical Education*, 69 :701–712, 1992. [19](#)
- [203] Z.Mihalić, S.Nikolić, and N.Trinajstić. Comparative study of molecular descriptors derived from the distance matrix. *J. Chem. InJ Comput. Sci.*, 32 :28–37, 1992. [87](#), [88](#)