Une méthode structurelle pour évaluer les propriétés d autosimilarité des plantes

Pascal Ferraro, Christophe Godin, Przemyslaw Prusinkiewicz, Lars Mundermann

▶ **To cite this version:**

Pascal Ferraro, Christophe Godin, Przemyslaw Prusinkiewicz, Lars Mundermann. Une méthode structurelle pour évaluer les propriétés d autosimilarité des plantes. Self Similarity And Applications Workshop, 2002, Clermont-Ferrand, France. hal-00306622

HAL Id: hal-00306622
https://hal.science/hal-00306622

Submitted on 25 Nov 2008

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

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

UNE MÉTHODE STRUCTURELLE POUR ÉVALUER LES PROPRIÉTÉS D'AUTOSIMILARITÉ DES PLANTES

PASCAL FERRARO[†*], CHRISTOPHE GODIN[†], PRZEMYSLAW PRUSINKIEWICZ[*] , AND LARS MÜNDERMANN[*]

Abstract. Dans de nombreux domaines de la biologie (arboriculture, sylviculture, botanique), la structure répétitive des plantes a pu être observée, mesurée et simulée. Toutefois, il n'existe pas de méthode pour l'évaluer quantitativement. Nous proposons dans cet article une méthode d'identidentification de caractères autosimilaires basée sur une comparaison structurelle des architectures de plantes. Cette méthode s'appuie sur un algorithme de comparaison entre arborescences, proposé en informatique théorique par Zhang. En utilisant cet algorithme, la distance entre deux systèmes ramifiés est définie comme le coût de la transformation de l'un en l'autre (à l'aide d'opérations élémentaires d'édition). Nous montrons dans ce papier comment cette distance peut alors être utilisée pour quantifier le degré d'autosimilarité de structures arborescentes. Nous illustrons cette méthodologie sur des inflorescences de lilas comme cas d'étude.

Key words. quantification, autosimilarité, structures arborescentes, plantes

1. Introduction. Le rôle joué par l'architecture dans la modélisation des plantes [13, 17, 19, 36] a généré le besoin de nouveaux outils d'investigation. Des outils génériques ont été développés pour visualiser l'architecture des plantes en 3-Dimensions [7, 33], pour modéliser le développement de la structure de la plante e.g. [21, 27, 8], pour mesurer son architecture [15, 18, 41], et pour explorer et analyser une plante [16]. Cet article introduit un nouvel outil pour l'identification des propriétés d'autosimilarité des plantes.

En 2001, Prusinkiewicz *et al.* [34] ont proposé une méthode permettant de modéliser de manière réaliste un ensemble de plantes, basée sur l'utilisation de leurs propriétés (supposées) d'autosimilarité. Ainsi, pour des besoins de visualisation, une inflorescence de lilas a été modélisée en utilisant uniquement les mesures effectuées sur le sommet d'une branche puis étendues à toute la structure (*cf.* figure 17a [34]). L'utilisation d'une telle modélisation nécessite d'évaluer au préalable la ressemblance des branches deux à deux, ou plus précisément d'évaluer un certain degré d'autosimilarité des inflorescences. Ferraro et Godin [11] ont proposé une distance entre architecture de plantes qui offre un outil efficace pour évaluer la ressemblance structurelle entre systèmes ramifiés. Dans cet article, nous développons une procédure, qui, partant des observations d'un système ramifié et utilisant cette distance, permet de quantifier le degré d'autosimilarité de la structure d'une plante et d'identifier les motifs identiques d'une structure arborescente.

Bien que la notion d'autosimilarité dans les plantes ne soit pas une notion botanique en tant que telle, ce concept peut être mis en relation avec les notions de "branch mapping", de relations paracladiales, d'axes de référence ou de réitérations développées dans la littérature. En nous appuyant sur une revue de ces concepts, (Section 2), nous développons les fondations mathématiques d'une distance entre architecture de systèmes ramifiés (Section 3) que nous utilisons pour formaliser le problème de la quantification de l'autosimilarité des plantes (Section 4). Nous illustrons enfin cette méthodologie sur des inflorescences de lilas comme un cas d'école. Nous terminons cet article par une discussion ouvrant les perspectives sur les travaux ultérieurs (Section 5).

2. Motivation Biologique. Pour de nombreux processus de croissance d'organismes vivant, et en particulier les plantes, la répétition régulière de certaines structures est aisément identifiable. Dans les cas les plus simples, les mêmes structures sont répétées périodiquement le long d'un axe, comme, par exemple les feuilles le long de la tige. Dans des cas plus complexes, la structure entière d'une étape précédente est répétée dans une partie de l'organisme à l'état suivant [26].

L'idée que les branches ont un comportement similaires aux branches qui les portent a fait l'objet de nombreuses études [13, 19, 9, 47, 2, 34]. Troll définit par exemple une inflorescence comme un système constitué d'une inflorescence principale et de *paracladia* [13], où un *paracladium* est une branche qui porte le même type de structure que celle portée dans la portion terminale de la branche mère [13, 23]. En s'appuyant sur cette définition, Fritjers et Lindenmayer ont introduit la notion de relation paracladiale entre deux systèmes ramifiés [13] : il y a relation paracladiale si "le comportement d'une branche est exactement le même que celui de la partie

[*]Department of Computer Science - University Of Calgary - Calgary, Alberta, T2N 1N4 Canada
[†]Plant Modelling Program - CIRAD TA 40/PS2 - Boulevard de la Lironde - 34398 Montpellier Cedex 5, France

de la branche porteuse qui s'est développée depuis la formation de la branche fille". Fritjers et Lindenmayer [13] puis Robinson [35] ont ainsi proposé une procédure permettant l'identification de formules de récurrence des relations paracladiales sur les formes ramifiées. Si ces concepts biologiques peuvent être observés visuellement dans des cas évidents, pour certaines plantes l'auto-similarité de la structure peut être cachée par la complexité de la structure. De tels développements peuvent donc être observés dans la nature mais rarement de manière exacte, dans tous les cas, ils n'ont été que très rarement évalué quantitativement.

De Reffye *et al.* [6] ont proposé une approche de modélisation de la croissance des plantes s'appuyant sur un principe similaire. Pour cela, ils ont introduit la notion *d'axe de référence* basée sur le regroupement et le classement de toutes les étapes de différentiation d'un arbre le long d'un axe théorique. L'étude de *Zelkova Serrata* (Orme du Japon) à différents stades de son développement a montré que la croissance et les processus de ramification étaient similaires dans toutes les parties de l'arbre. Ils ont ainsi statistiquement identifié un axe théorique qui, de la graine à la mort de l'arbre, donne tous les états possible de la différentiation morphologique d'une unité de croissance[1]. Dans ce sens, "l'axe de référence" interprète le comportement des bourgeons dans une plante en s'appuyant sur une analyse quantitative des gradients des tailles d'organes sur les différents axes.

Nous proposons dans ce qui suit un outil permettant d'évaluer automatiquement la ressemblance entre différentes structures ramifiées dans une plante et ainsi d'identifier les relations paracladiales ou les structures homologues à la base du modèle d'axe de référence. Cet outil est basé sur une représentation particulière des sytèmes ramifiés à l'aide de graphes et s'appuie sur une comparaison structurelle de ces représentations.

3. Évaluation de la ressemblance entre deux systèmes ramifiés. Pour comparer deux systèmes ramifiés, une première approche consiste à résumer chaque individu par un petit nombre de variables synthétiques et globales (*e.g.* production de fruits, taille de la couronne, *etc.*). L'évaluation de la ressemblance de deux systèmes ramifiés est alors réduite à la comparaison ces variables synthétiques. Ceci définit les méthodes de comparaison globales pour lesquelles l'organisation topologique des entités dans la structure ramifiée n'est pas prise en compte. Il existe cependant des domaines pour lesquels la structure d'un système ramifié joue un rôle important. Dans de tels cas, la notion de distance entre deux systèmes ramifiés devrait naturellement prendre en compte l'organisation spatiale des entités. Ceci définit les méthodes de comparaison analytique qui sont basées sur une comparaison élément par élément [28].

3.1. Définitions. Une plante est considérée comme un ensemble d'entités botaniques (*e.g.* entre-noeuds et noeuds, unités de croissance, pousses annuelles). L'organisation topologique peut être représentée par un graphe [14]. Un graphe $G = (V, E)$ est constitué d'un ensemble V de vertex et d'un ensemble E d'arcs, chaque arc étant représenté par une paire de vertex [32]. Les vertex représentent les entités botaniques et les arcs correspondent aux connections physiques entre celles-ci. Si (v_1, v_2) est un arc de E, le vertex v_1 est appelé père de v_2 et le vertex v_2 est appelé le fils de v_1 [32]. L'ensemble des fils de v_1 est noté $son[v_1]$. Dans un graphe, un *chemin* de x_1 à x_n est une séquence de vertex $(x_1, x_2.., x_n)$ telle que pour deux vertex consécutifs $\{x_i, x_{i+1}\}$ de la séquence, (x_i, x_{i+1}) est un arc. Dans un graphe, un vertex v_1 est appelé l'*ancêtre* d'un autre vertex s'il existe un chemin de v_1 à v_2.

On affecte à chaque vertex un ou plusieurs attributs qui représentent des caractéristiques biologiques des entités. Chaque attribut est une valeur numérique réelle (*e.g.* le diamètre ou la longueur d'une entité) ou symbolique (*e.g.* le type d'une entité). Ces attributs seront donc considérés comme choisis dans un ensemble de label. Soit $\Sigma = a, b, c, \ldots$ un ensemble fini ou infini de labels, et soit λ un unique symbole dans Σ. Soit α une fonction d'étiquetage qui affecte un label à tout vertex d'un graphe. Nous suposons qu'une distance est d, appelée distance élémentaire, est définie sur l'ensemble des labels Σ. Une distance entre deux vertex v_1 et v_2 d'un graphe est alors définie en utilisant la distance élémentaire sur Σ : $d(v_1, v_2) = d(\alpha(v_1), \alpha(v_2))$. Nous supposons de plus, qu'un coût $\gamma(x)$ est affecté à tout label x tel que :

$$\forall (x, y) \in \Sigma : d(x, y) \leq \gamma(x) + \gamma(y)$$

Un coût sera de plus affecté à chaque vertex v : $\gamma(v) = \gamma(\alpha(v))$.

Dans une plante $\mathcal{P}$, toute entité étant physiquement attachée à au plus une entité parente, la structure topologique est représentée par une arborescence [14] $T = (V, E)$. Une arborescence est un graphe dans lequel

[1] portion de la tige mise en place durant une période d'allongement continue [4]

tout vertex excepté un seul, r, appelé *racine*, a un et un seul parent. La racine n'a pas de parent. Dans la suite une arborescente enracinée en r sera représentée par $T[r]$. Afin d'identifier les différents axes (*cf.* section 4) d'une plante représentée par une arbroescence $T[r]$, les arcs de l'arborescence sont étiquetés : une entité peut soit précéder (symbôle '<') soit porter (symbôle '+') une autre entité. Nous supposerons par la suite qu'une entité ne peut être connectée qu'à au plus une entité par un arc '<'. Une forêt est un graphe dont les composantes connexes sont des arbres. Pour tout vertex x d'une arborescence $T[v]$, $F[x]$ représente la *forêt enracinée en* x, *ie.* la forêt obtenue en enlevant à l'arborescence $T[x]$ sa racine x et tout les arcs incident de x. Dans la suite nous utiliserons le terme *forêt* pour désigner une forêt enracinée. L'ensemble de tous les sous-arbres et forêts seront respectivement notés $\mathcal{T}$ et $\mathcal{F}$.

3.2. Comparaison d'arborescences . La représentation des données sous forme d'arborescences totalement étiquetées permet d'exprimer une information structurelle, soit de type concret (géométrique, dans le cas d'images, par exemple), soit de type abstrait (relations hiérarchiques entre les éléments primitifs). La manipulation de ces structures nécessite la définition d'algorithmes efficaces. En reconnaissance de formes, la nécessité de comparer deux objets représentés par des arborescences (*e.g.* reconnaissance de caractères [24], corrélation de signaux [5], sciences du comportement [3], classification hiérarchique [25], chimie [43], biologie moléculaire [39], ...) pose fréquemment le problème de la comparaison de deux arborescences.

On peut distinguer deux approches spécifiques permetant de définir une mesure de dissimilarité entre deux arborescences. Une première approche, développée par des mathématiciens algébristes, définit des métriques sur des ensembles partiellement ordonnés (dont les arborescences sont un cas particulier). Ces travaux concernent généralement les arborescences dont seules les feuilles sont étiquetées, et sont appliqués en particulier à la comparaison des résultats d'algorithmes de classification hiérarchique sur un ensemble de données [3, 29, 20]. Une deuxième approche, communément appelée *problème de l'édition d'arborescences* (*tree editing problem* [38], en anglais), consiste à effectuer une suite de transformations permettant de construire une arborescence cible à partir d'une arborescence initiale. La classification hiérarchique [25], la reconnaissance de formes [28, 37, 1] l'analyse de scène [46], la chimie [43] ou la biologie moléculaire [39, 40] sont autant de domaines d'applications du calcul d'une distance d'édition. Ce calcul est basé sur la définition de trois opérations l'insertion, l'élision et la substitution, d'autres opérations pouvant être envisagées, comme la fusion ou l'éclatement de vertex [37]. Ces méthodes font apparaître plusieurs stratégies spécifiques qui s'appuient sur des définitions différentes des opérations d'édition. Selkow [38] a proposé dans un premier temps d'appliquer les opérations d'insertion et d'élision uniquement aux feuilles des arborescences. Taï [42] définit une métrique pour laquelle l'insertion et l'élision de vertex s'appliquent à tout vertex des arborescences concernées mais qui conserve un ordre partiel prédéfini sur les vertex des arborescences (*arborescences ordonnées*). Entre ces deux métriques, de nombreuses autres distances ont été définies [25, 31, 44, 45] dérivant toutes de la distance de Taï [37]. Celle de Whilhelm [48], par exemple, considère une extension des distances de Taï qui s'affranchit de la conservation de la numérotation des vertex pour des arborescences ordonnées dont les étiquettes sont toutes deux à deux distinctes. Enfin, dans le cas d'arborescences non-ordonnées, Zhang [49, 50] propose une modification des contraintes de Taï permettant de définir une nouvelle distance entre arborescences. Les trois opérations d'édition considérées sur toute arborescence non ordonnée sont les suivantes :

- la substitution d'un vertex v_1 par un vertex v_2 (cette opération consiste à modifier les labels des deux vertex v_1 et v_2);
- l'élision d'un vertex v , *ie.* l'ensemble des fils de v deviennent les fils du père v et v est effacé de l'arborescence T (ainsi que les arcs incidents de v);
- l'insertion d'un vertex v comme fils d'un vertex x : *ie.* l'opération symmétrique de l'élision, l'ensemble des fils de x devient l'ensemble des fils de v et v devient l'unique fils de x.

Les différentes opérations d'insertion et d'élision sont représentées sur la Figure 3.1. Notons qu'une opération de substitution est une fonction de l'ensemble des arborescences sur lui même alors qu'une fonction d'élision est une fonction de l'ensemble des arborescences $\mathcal{T}$ vers l'ensemble des forêts $\mathcal{F}$ (avec la relation $\mathcal{T} \subset \mathcal{F}$). L'ensemble des opérations d'édition possible est notée $\mathcal{E}$.

Pour toute opération d'édition s, un coût $\gamma(s)$ est défini comme suit :

- $\gamma(s) = \gamma(C(x,y)) = d(x,y)$ si s est une opération $C(x,y)$ qui substitue les vertex x et y;
- $\gamma(s) = d(D(x)) = \gamma(x,\lambda)$ si s est une opération $D(x)$ qui efface le vertex x et,

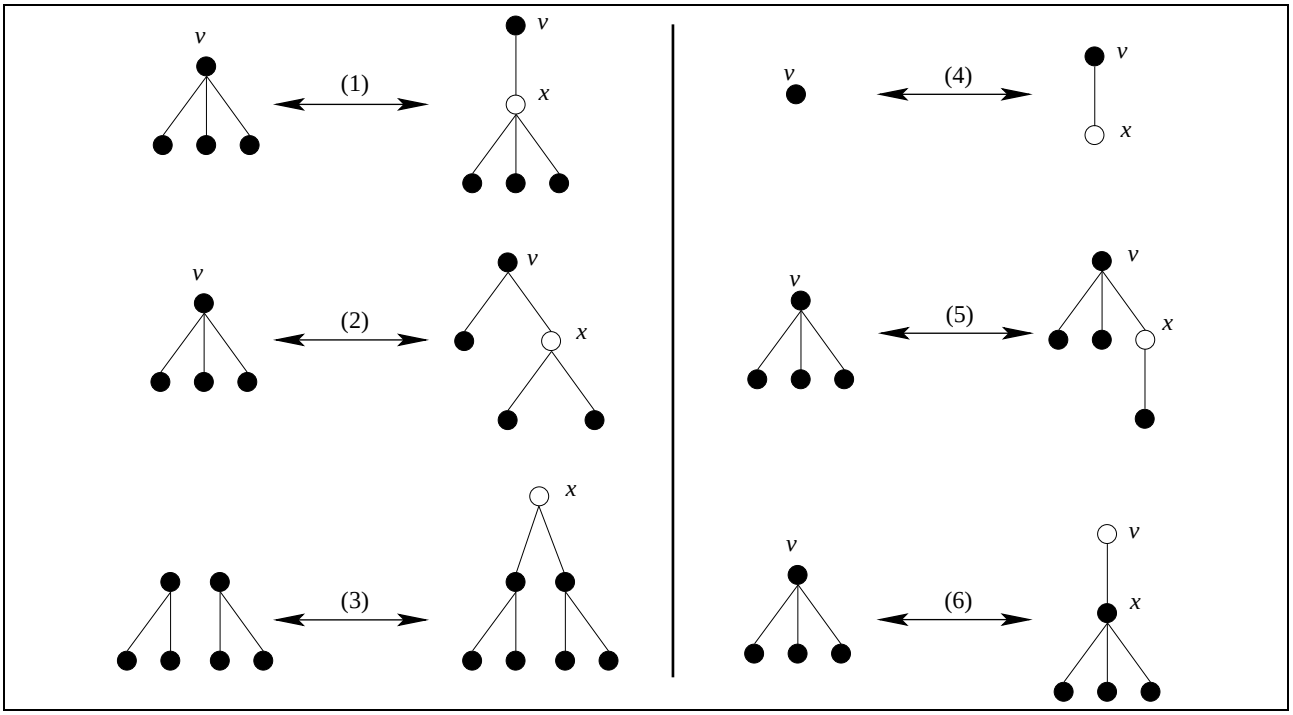

FIG. 3.1. *Les différentes opérations d'insertion et d'élision possible sur une arborescence. Les flèches de gauche à droite représentent une opération d'insertion, et les flèches de droite à gauche représentent les opérations d'élision. (1) L'ensemble des fils de v devient l'ensemble des fils de x, (2) un sous-ensemble strict des fils de v devient l'ensemble des fils de x, (3) x est insérée comme racine d'un ensemble d'arbres. Les opérations (4) à (6) sont des restrictions des opérations (1) et (3) mais ce sont des opérations rencontrées fréquemment. Dans l'algorithme, les opérations de type (2) ne sont pas autorisées.*

- $\gamma(s) = d(I(y)) = \gamma(\lambda, y)$ si s est $I(y)$ une opération qui insère le vertex y.

Nous rappelons que d et γ vérifie la propriété suivante :

$$d(x,y) \leq \gamma(x) + \gamma(y) \tag{3.1}$$

Nous considérons deux arborescences $T_1[v]$ et $T_2[w]$, nous noterons $\mathcal{S}(v,w)$ l'ensemble de toutes les séquences d'opérations d'édition qui transforme $T_1[v]$ en $T_2[w]$. Soit S une séquence de n opérations d'édition $(s_1, s_2, \ldots, s_n)$ de $\mathcal{S}(v,w)$, le coût $\gamma(S)$ de la séquence est défini en sommant le coût des différentes opérations qui composent la séquence : $\gamma(S) = \sum_{s_i \in S} \gamma(s_i)$. La mesure de ressemblance $D(T_1[v], T_2[w])$ entre les deux arborescences est définie comme la séquence de $\mathcal{S}(v,w)$ de coût minimum :

$$D(T_1, T_2) = \min_{S \in \mathcal{S}} \{\gamma(S)\} \tag{3.2}$$

Il est possible de montrer [12] que la mesure de ressemblance ainsi définie est égale à la distance entre deux arborescences définie par Zhang [50]. Cet auteur a proposé un algorithme de programmation dynamique [49, 50] calculant la distance entre deux arborescences T_1 et T_2 en un temps :

$$O\left(|T_1| \times |T_2| \times \max\left(\deg(T_1) + \deg(T_2)\right) \times \log_2\left(\deg(T_1) + \deg(T_2)\right)\right)$$

Où $\deg(T_1)$ (resp.$\deg(T_2)$ représente le nombre maximal de fils de l'arborescence T_1 (resp. T_2). Cet algorithme [49, 50] s'appuie sur un calcul récursif des distances entre les sous-arbroescences de $T_1[v]$ et $T_2[w]$ (détaillé dans [10]) :

$$D(T_1[v], T_2[w]) = \min \begin{cases} D(\theta, T_2[w]) + \min_{w_k \in son[w]} \{D(T_1[v], T_2[w_k]) - D(\theta, T_2[w_k])\} \\ D(T_1[v], \theta) + \min_{v_k \in son[v]} \{D(T_1[v_k], T_2[w]) - D(T_1[v_k], \theta)\} \\ D(F_1[v], F_2[w]) + d(v,w) \end{cases}$$

Ce calcul fait apparaître le calcul de la distance entre les deux forêts $F_1[v]$ et $F_2[w]$ issues de v et w. On montre de la même façon que ce calcul s'effectue récursivement et s'appuie lui même récursivement sur le calcul de la distance entre arborescences [49, 50].

Algorithmiquement, le calcul de la distance $D(T_1[v], T_2[w])$ nécessite celui d'une matrice M de dimension $n \times p$, où n et p sont respectivement les nombres de vertex de $T_1[v]$ et $T_2[w]$. Si on donne un ordre arbitraire (préfixé par exemple) sur les vertex des deux arborescences, le coefficient $m_{i,j}$ de M représente alors la distance $D(T_1[x], T_2[y])$ entre les sous-arborescences $T_1[x]$ et $T_2[y]$, où x et y sont respectivement les vertex numéro i et j des arborescences $T_1[v]$ et $T_2[w]$, la distance entre les deux arborescences $T_1[v]$ et $T_2[w]$ étant obtenue au coefficient $m_{0,0}$ de M (avec l'ordre préfixé).

3.3. Standardisation et normalisation de la distance obtenue. La distance entre arborescences basée sur une transformation d'édition, s'appuie sur la définition d'une distance élémentaire entre les vertex des arborescences. Cette distance affecte un réel positif $d(x_1, x_2)$ à toute paire de vertex (x_1, x_2) des deux arborescences $T_1[v]$ et $T_2[w]$ et un coût, réel positif $\gamma(x)$ à tout vertex x, vérifiant :

$$\forall (x, y) \in T_1[v] \times T_2[w] : d(x, y) \leq \gamma(x) + \gamma(y)$$

Une première distance élémentaire est la *distance de Levenstein* [22]. C'est une fonction de coût local élémentaire basée sur une distance binaire. Dans ce cas, un coût nul est affecté à toute opération de substitution, et un coût de un pour toute opération d'insertion ou d'élision. Le coût local ainsi défini ne tient pas compte de la nature des vertex. La distance d'édition qui en découle est donc indépendante des vertex concernées par l'opération. Une distance basée sur cette fonction de coût local tient compte uniquement de la topologie des structures comparées, nous l'appellerons *distance topologique*. Le coût binaire ainsi défini peut être raffiné en utilisant les attributs liés aux entités tels que la longueur, le diamètre de la moëlle, le type, etc. Nous étudions dans ce qui suit ce coût variable en séparant les coûts de substitution de ceux de l'élision et de l'insertion. Dans le cas où l'on considère plusieurs attributs numériques ou symboliques [11], il est nécessaire d'homogénéiser la dynamique des différents attributs afin qu'ils aient une importance comparable dans la définition de la distance. Une procédure de standardisation est alors nécessaire.

Nous avons montré dans [11] que, dans le cas d'une distance locale topologique, la distance maximale entre deux arborescences est égale à la somme des tailles (du nombre de vertex) des deux arborescences comparées. La distance obtenue par l'algorithme de Zhang sera donc normalisée de la manière suivante :

$$D'(T_1[v], T_2[w]) = \frac{D(T_1[v], T_2[w])}{(|T_1[v]| + |T_2[w]|)} \tag{3.3}$$

La distance ainsi obtenue est nécessairement comprise entre 0. et 1. et s'affranchit au mieux de la différence des tailles des deux arborescences.

4. Evaluation de l'autosimilarité dans une arborescence. Afin d'évaluer l'autosimilarité au sein d'une arborescence, nous proposons d'appliquer la méthode de comparaison structurelle, présentée précédemment à une même arborescence. En effet, le calcul de la distance entre une arborescence $T[v]$ et elle même selon l'algorithme de Zhang nécessite le calcul de la distance de l'ensemble des sous-arborescences de $T[v]$ deux à deux. En comparant $T[v]$ avec elle-même, nous faisons apparaître le calcul d'une matrice de distance M, de diagonale nulle et dont les coefficient $m_{i,j}$ sont les valeurs de la distance entre les sous-arborescences $T[v_i]$ et $T[v_j]$, où v_i et v_j sont deux vertex de $T[v]$. L'étude de cette matrice nous permet de mettre alors en évidence les sous-arborescences de $T[v]$ homologues.

De manière générale, pour l'évaluation de l'autosimilarité d'une arborescence, nous appliquerons des méthodes de classification sur cette matrice de distance. Dans le cas des plantes (représentées par une arborescence), nous émettons l'hypothèse qu'une plante reproduit sur ces ramifications toute (ou partiellement) la structure de la partie sommitale de la plante. Dans la suite de cette section, nous introduisons les définitions mathématiques permettant de vérifier une telle hypothèse.

4.1. Définitions. Dans ce qui suit, nous considérons une plante $\mathcal{P}$ représentée par un graphe $T[r]$. L'ensemble de toutes les sous-arborescences de $T[r]$ est noté $\mathcal{T}(r)$.

Soit v un vertex de $T[r]$ et soit $T[v]$ une sous-arborescence enracinée en v, nous définirons alors *l'axe* $A[v] = (V, E)$ de $T[v]$ comme le sous-graphe de $T[v]$ tel que $V = \{v_1, v_2, \ldots, v_n\}$ où $v = v_1$ et, pour tout $i \in \{1 \ldots n-1\}$, (v_i, v_{i+1}) est un arc de type '<' (figure 4.1). La hauteur de l'arborescence $T[v]$ est définie comme le nombre d'éléments de l'axe $A[v]$.

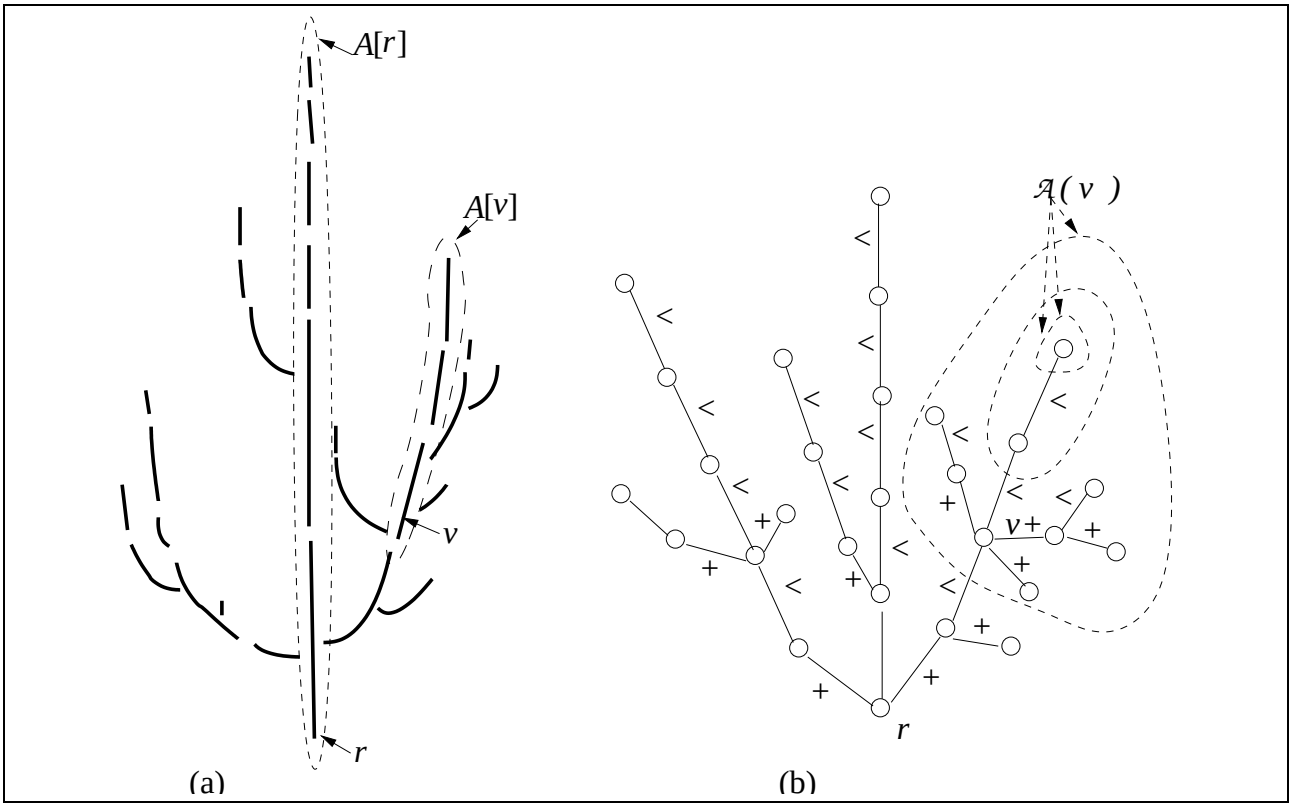

FIG. 4.1. *(a) un système ramifié et (b) sa représentation sous forme d'arborescence.*

L'ensemble des sous arborescences de $T[r]$ enracinée en un vertex de $A[v]$ est noté $\mathcal{A}(v)$:

$$\mathcal{A}(v) = \{T[v_1], T[v_2], T[v_3], \ldots, T[v_n]\}$$

Remarquons que dans cette définition v_{i+1} est nécessairement un fils de v_i (pour tout i dans $\{v_1, \ldots, v_{n-1}\}$). De plus, $\mathcal{A}(v)$ est un ensemble d'arborescences emboîtées les unes dans les autres (*cf.* figure 4.1). L'axe $A[r]$ est appelé *axe principal* de $\mathcal{P}$.

En nous appuyant sur la définition de la notion de *relation paracladiale* [13, 23], nous donnons une évaluation de l'autosimilarité en identifiant la ressemblance des différentes structures arborescentes avec la tige principale :

DEFINITION 1. *Pour toute sous-arborescence $T[v]$ de $\mathcal{T}(r)$, le* coefficient paracladial *de $T[v]$ est la quantité :*

$$pr(v) = \min_{r_i \in A[r]} \{D(T[v], T[r_i]))\}$$

Pour une ramification quelconque $T[v]$, ce coefficient correspond à l'évaluation de la ressemblance entre la branche et la portion sommitale de l'axe principal qui est la plus ressemblante avec la branche $T[v]$. Nous dirons qu'une plante, représentée par une arborescence $T[r]$ est "*autosimilaire*" si toute sous-arborescence $T[v]$ de $\mathcal{T}(r)$ vérifie :

$$\forall T[v] \in \mathcal{T}(r) \qquad pr(v) = 0$$

En général dans la nature on ne rencontre pas de système ramifié vérifiant parfaitement des propriétés d'autosimilarité. Dans des applications pratiques, pour évaluer de manière quantitative ces notions, nous considérons la valeur moyenne sur toutes les sous-arborescences $T[v]$ de $\mathcal{T}(r)$ du coefficient paracladial :

$$\overline{PR} = \frac{1}{|\mathcal{T}(r)|} \times \sum_{T[u] \in \mathcal{T}(r)} D(T[u], T[r_u]) \tag{4.1}$$

où $|\mathcal{T}(r)|$ représente le nombre d'arborescences dans $\mathcal{T}(r)$ (*ie.* le nombre de vertex de $T[r]$)

4.2. Étude de la valeur de $\overline{PR}$. Les valeurs de $\overline{PR}$ sont entre 0.0 et 1.0, la valeur maximale (1.0) étant atteinte lorsqu'on compare une arborescence avec un arbre nul (sans vertex, ni arc). Lorsque $\overline{PR}$ a une valeur proche de 0.0 on peut alors estimer que la structure est "quasi-autosimilaire".

D'un point de vue pratique, pour évaluer l'auto-similarité , nous devons tout d'abord définir une distance locale entre les vertex représentant les entités élémentaires de la plante. Dans nos exemples, nous n'utiliserons qu'une distance topologique, *ie.* nous ne tiendrons pas compte des éventuels attributs numériques ou qualitatifs affectés aux vertex. Dans ce cas là, la valeur de $\overline{PR}$ correspond au pourcentage moyen de vertex qui ont été insérés ou supprimés durant la comparaison. Un faible pourcentage indique alors que les ramifications sont très ressemblantes à la tige principale, et donc qu'il y a une forte autosimilarité de la structure.

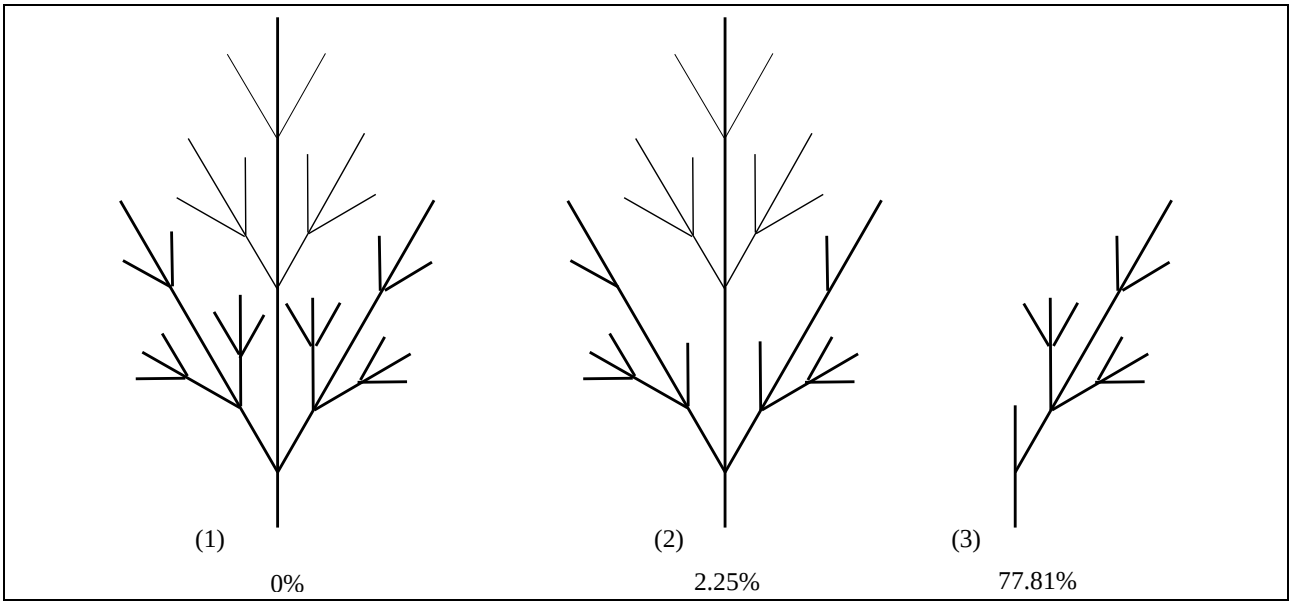

FIG. 4.2. *Trois structures ramifiées (a) parfaitement autosimilaire, (b) partiellement autosimilaire et (c) ne présentant pas un caractère autosimilaire. Le pourcentage indiqué est la valeur de $\overline{PR}$ dans chacun des cas.*

Considérons par exemple les deux structures ramifiées (a), (b) et (c) de la figure 4.2. Pour la structure ramifiée (a), la structure parfaitement autosimilaire nous permet d'obtenir un coefficent paracladial nul. On peut constater que ce coefficient croit légérèrement lorsqu'on dégarde cette structure d'origine de quelques vertex, mais augmentre très sensiblemen lorsque la structure est fortement déséquilibrée (grande ramification portée par un tronc réduit à son minimum).

4.3. Analyse d'inflorescences de lilas. Nous présentons dans ce qui suit une application dans un cas réel de l'approche présentée ci-dessus. Dans cet exemple, nous considérons 5 inflorescences de lilas (noté A_1 à A_5), choisi pour leurs propriétés d'auto-similarité. Nous cherchons à évaluer quantitativement ces propriétés qui peuvent être observé visuellement. Les systèmes ramifiés représentant 5 inflorescences de lilas sont représentées sous la forme d'une arborescence, chaque vertex de l'arborescence représente un "entre-nœud" (portion d'axe entre deux ramifications ou fleurs). Les longueur et diamètre de chaque entre-nœud ont été mesurés.

D'un point de vue pratique, pour évaluer l'auto-similarité , nous devons tout d'abord définir une distance locale entre entités élémentaires (ici les entre-nœuds) des inflorescences. Dans notre exemple, nous n'utiliserons

qu'une distance topologique, *ie.* nous ne tiendrons pas compte des mesures réalisées sur les entre-nœuds. Dans ce cas là, la valeur de $\overline{PR}$ correspond au pourcentage moyen de nœuds qui ont été insérés ou supprimés durant la comparaison. Un faible pourcentage indique alors que les ramifications sont très ressemblantes à la tige principale, et donc il y a une forte autosimilarité de la structure.

Le tableau 4.1 indique les valeurs de $\overline{PR}$ pour chaque inflorescence, indiquant pour chacune d'elle un certain degré d'autosimilarité.

	(a) $\overline{PR}$	(b)
A_1	1.64%	70.83%
A_2	1.96%	78.26%
A_3	2.12%	68.18%
A_4	2.27%	79.16%
A_5	1.92%	72.72%

TABLE 4.1

(a) Valeur de $\overline{PR}$ pour chaque inflorescence, et (b) pourcentage de branches d'ordre 2 ayant un axe de même taille que la structure homologue identifiée par l'algorithme sur la tige principale.

Nous pouvons remarquer que pour chaque inflorescence, les valeurs de $\overline{PR}$ sont proches de 0. Pour mettre en évidence cette remarque, la distance entre entre la branche et la portion la plus ressemblante sur la tige principale peut être observée graphiquement en utilisant une reconstruction 3-D des inflorescences [30]. La figure 4.3 indique la valeur du cœfficient paracladial pour chaque branche. On peut remarquer que les plus fortes valeurs de $pr(v)$ sont généralement obtenues pour les branches du sommet. Les branches à la base de l'inflorescence qui ont eu une croissance plus importante et donc plus de chance de "subir" des accidents extérieur. On peut également observer une certaine variablilité du cœfficient paracladial au sein de la population, l'inflorescence A_1 étant visiblement poss'édant le meilleur taux de répétition dans sa structure.

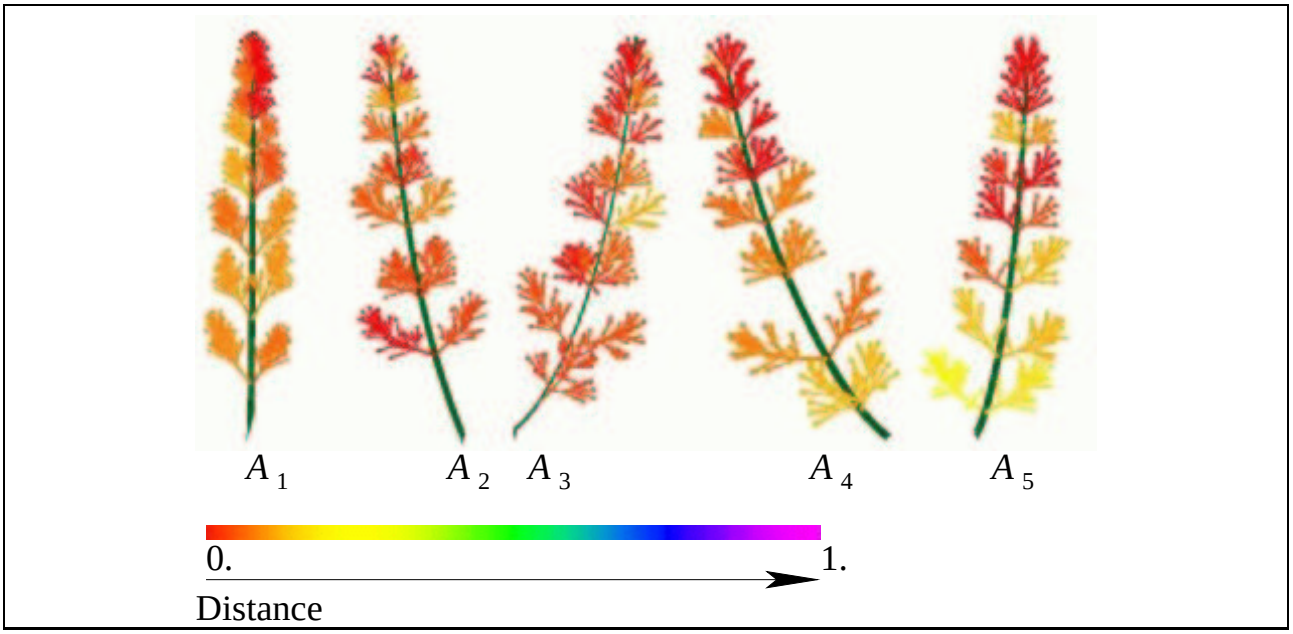

FIG. 4.3. *Représentation du coefficient paracladial pour chaque branche de chaque inflorescence.*

La méthode présentée précédemment permet de plus d'identifier quelles sont les sous- parties homologues à l'intérieur de l'inflorescence. Si nous supposons qu'il existe une relation entre la taille de l'axe principal et une branche, cette supposition peut être vérifiée en eutilisant une telle analyse. Le tableau 4.1 donne le nombre de branche d'ordre 2 qui ont exactement la même taille que la structure homologue identifiée par l'algorithme

sur la tige principale et indique une "bonne" corrélation entre le nombre d'entités de l'axe de référence et une branche d'ordre 2.

5. Conclusion. Dans cet article, nous avons proposé une méthode analytique d'évaluation de l'autosimilarité des plantes. La quantification de l'autosimilarité dans une plante s'effectue à partir d'une comparaison structurelle des différents sous-systèmes ramifiés de la plante avec la tige principale. La distance entre deux structures ramifiées, fondée sur une méthode d'édition, nécessite le calcul d'une matrice de distances des sous-systèmes ramifiés deux à deux de la plante. La méthode propose une analyse d'une partie de la matrice, plus précisément, celle correspondant aux cœfficients de la tige principale. L'utilisation partielle de cette matrice nous permet tout de même l'identification de propriétés d'autosimilarité chez certaine plante autorisant ainsi la réduction du nombre de mesures nécessaires à la construction de modèles (*cf.* [34]). Néanmoins, pour une évaluation plus complète de l'autosimilarité, les développements futurs de cette méthode prendront en compte l'ensemble de la matrice, c'est à dire la comparaison de toutes les sous-parties de plantes deux à deux.

L'appproche que nous proposons ici est une étape dans la l'évaluation quantitative de la production de structures équivalentes et emboitées chez les plantes. Nous envisageons par exemple d'utiliser cette méthode et ses extensions pour automatiser (au moins partiellement) l'inférence de modèles de croissance de plantes.

REFERENCES

[1] F. Banegas, D. Michelucci, M. Roelens, and M. Jaeger. Significant features extraction from 3d points cloudsby ellipsoidal skeletons. application in vision and geometric characterization. In *Internationnal Conference in Visual Computing*, 1999.

[2] D. Barthélémy. Levels of organization and repetition phenomena in seed plants. *Acta Biotheoretica*, 39:309–323, 1991.

[3] S. Boorman and D. Oliver. Metrics on spaces of finite trees. *J. Math. Psychol*, 10(26-59), 1973.

[4] Y. Caraglio and D. Barthélémy. Revue critique des termes relatifs à la croissance et à la ramification des tiges des végétaux vasculaires. In I. édition, editor, *Modélisation et simulation de l'architecture des végétaux*, chapter Part I, pages 11–87. Science Update, 1997.

[5] Y. Cheng and S. Lu. Waveform correlation by tree matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 7:299–395, 1985.

[6] P. de Reffye, P. Dinouard, and D. Barthélémy. Architecture et modélisation de l'orme du japon zelkova serrata (thunb.) makino (ulmaceae): la notion d'axe de référence. In *De la forêt cultivée à l'industrie de demain: 3e colloque Sciences et Industries du Bois, Arbora*, pages 351–352, Bordeaux, France, 1990.

[7] P. de Reffye, E. Elguero, and E. Costes. Growth units construction in trees: a stochastic approach. *Acta Biotheoretica*, 39:325–342, 1989.

[8] P. deReffye, T. Fourcaud, F. Blaise, D. Barthélémy, and F. Houllier. A functional model of tree growth and tree architecture. *Silva Fennica*, 31(3):297–311, 1997.

[9] C. Edelin. *L'architecture monopodiale : l'exemple de quelques arbres d'asie tropicale.* PhD thesis, USTL, Montpellier II, 1984.

[10] P. Ferraro. *Méthodes algorithmiques de comparaison d'arborescences. Applications à la comparaison de l'architecture des plantes.* Informatique, Institut National Polytechnique de Toulouse (France), Toulouse (France), November 2000.

[11] P. Ferraro and C. Godin. A distance measure between plant architectures. *Annals of Forest Science*, 57:445–461, june 2000.

[12] P. Ferraro and C. Godin. An unified point of view on tree graph comparison. *Algorithmica*, Submission March 2002.

[13] D. Frijters and A. Lindenmayer. Developmental descriptions of branching patterns with paracladial relationships. *Automata, Langauges, Development*, 1976.

[14] C. Godin and Y. Caraglio. A multiscale model of plant topological structures. *Journal of theoretical biology*, 191:1–46, 1998.

[15] C. Godin, E. Costes, and Y. Caraglio. Exploring plant topological structure with the amapmod software: an outline. *Silva Fennica*, 31:355–366, 1997.

[16] C. Godin, Y. Guédon, E. Costes, and Y. Caraglio. Measuring and analyzing plants with the amapmod software. In M. Michalewicz, editor, *Advances in computational life sciences, Vol I : Plants to ecosystems*, volume January, pages 63–94. Csiro, Australia, 1997. chapitre 4.

[17] F. Hallé, R. Oldeman, and P. Tomlinson. *Tropical trees and forests. An architectural analysis.* Springer Verlag, New York, 1978.

[18] J. S. Hanan and P. Room. Practical aspects of plant research. In M. Michalewicz, editor, *Advances in computational life sciences, Vol I : Plants to Ecosystems*, volume January, pages 28–43. CSIRO, Australia, 1997. Chapitre 2.

[19] H. Honda. Description of the form of trees by the parameters of tree-like body. *Journal of theoretical biology*, 1971.

[20] M. Karonski and Z. Palska. On ms type distances between hypergraphs. *Zatosowania Matermatyki*, XVI.1, 1977.

[21] W. Kurt. Growth grammar interpreter grogra 2.4: A software for the 3-dimentional interpretation of stochastic, sensitive growth grammar in the context of plant modelling. Introduction and reference manual, Forschungszentrum Waldokosysteme der Universitat Gottingen, 1994.

[22] A. Levenstein. Binary codes capable of correcting deletions, insertions and reversals. *Sov. Phy. Dohl.*, 10:707–710, 1966.

[23] A. Lindenmayer. Paracladial relationships in leaves. *Ber. Deutsch. Bot. Ges. Bd*, 90:287–301, 1977.

[24] S. Lu and K. FU. Error-correcting tree automata for syntactic pattern recognition. *IEEE Trans. Computers*, 27:1040–1053, 1978.

[25] S.-Y. Lu. A tree-to-tree distance and its application to cluster analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1:219–224, 1979.
[26] G. M. Herman, A. Lindenmayer, and G. Rozenberg. Description of developmental languages using recurence system. *Mathematical systems theory*, 8(4):316–341, 1975.
[27] R. Mech and P. Prusinkiewicz. Visual models of plants interacting with their environment. In H. Rushmeier, editor, *SIGGRAPH 96 Conference Proceedings*, pages 397–410, New Orleans (Louisiana, USA), 1996. Addison-Wesley.
[28] L. Miclet. *Méthodes Structurelles Pour la Reconnaissance Des Formes*. Eyrolles, 61, bd Saint-Germain - 75005 Paris, 1984.
[29] B. Monjardet. Metrics on partially ordered sets, a review. *Discr. Math.*, 35:173–184, 1981.
[30] R. Měch. CPFG version 3.4 user's manual. Technical report, 1998. Department of Computer Science, University of Calgary.
[31] K. Nakabayashi and K. Kamata. A distance between two binary trees and its computing algorithm. *Tans. IECE*, J66-D(4):445–451, 1983.
[32] F. P. Preparata and R. T. Yeh. *Introduction to Discrete Structures for Computer Science and Engineering*. Addison-Wesley Publishing Company, 1973.
[33] P. Prusinkiewicz and A. Lindenmayer. *The algorithmic beauty of plants*. Springer Verlag, New York, 1990.
[34] P. Prusinkiewicz, L. Mündermann, R. Karwowski, and B. Lane. The use of positional information in the modeling of plants. In A. SIGGRAPH, editor, *Computer Graphics Proceedings, Annual Conference Series, 2001*, pages 289 – 300, Los Angeles, CA, August 2001.
[35] D. F. Robinson. A notation for the growth of inflorescences. *The New Phytologist*, 53:587–596, 1986.
[36] J. Ross. *The radiation regim and the architecture of plant stands*. Junk W. Pubs., The Hague, The Netherlands, 1981.
[37] A. Sanfeliu. Matching tree structures. In H. Bunke and A. Sanfeliu, editors, *Syntactic and Structural pattern recognition. Theory and applications*, volume 7 of *Series in Computer Science*, pages 144–178. Word Scientific, 1990.
[38] S. M. Selkow. The tree-to-tree editing problem. *Information processing letters*, pages 184–186, 1977.
[39] B. A. Shapiro and K. Zhang. Comparing multiple rna secondary structures using trees comparisons. *Cabios*, 6:309–318, 1990.
[40] D. Shasha, J. T.-L. Wang, K. Zhang, and F. Y. Shih. Exact and approximate algorithms for unordered tree matching. *IEEE transactions on systems, man, and cybernetics*, 24(4):668–678, April 1994.
[41] H. Sinoquet, P. Rivet, and C. Godin. Assessment of the three-dimensional architecture of walnut trees using digitizing. *Silva Fennica*, 3:265–273, 1997.
[42] K.-C. Tai. The tree-to-tree correction problem. *Journal of the Association for Computing Machinery*, pages 422–433, 1979.
[43] Y. Takajashi, Y. Satoh, H. Suzuki, and S. Saski. Recognition of largest common structural fragment among a variety of chemical structures. *Anal. Sci*, 3:23–28, 1987.
[44] E. Tanaka. The metric between trees based on the strongly structure preserving mapping and its computing method. *Trans. IECE*, J67-D(6):722–723, 1984.
[45] E. Tanaka. A computing algorithm for the tree metric based on the structure preserving mapping. *Trans. IECE*, E68(5):317–324, 1985.
[46] J. Wernecke. *The inventor Mentor. Programming Object-Oriented 3D Graphics with Open Inventor, Release 2*. Addison Wesley, 1995.
[47] J. White. The plant as a metapopulation. *Ann. Rev. Ecol. Syst.*, 10:109–145, 1979.
[48] R. Wilhelm. A modified tree-to-tree correction problem. *Information processing letters*, 12(3):127–132, June 1981.
[49] K. Zhang. A new editing-based distance between unordered trees. In *Combinatorial Pattern Matching, 4th Ann. Symp.*, pages 254–265, Padala (Italy), 1993. CPM'93.
[50] K. Zhang. A constrained edit distance between unordered labeled trees. *Algorithmica*, 15:205–222, 1996.