Treelet Kernel Incorporating Chiral Information

Pierre-Anthony Grenier, Luc Brun, Didier Villemin

Treelet Kernel Incorporating Chiral Information

Pierre-Anthony Grenier†, Luc Brun†, and Didier Villemin‡

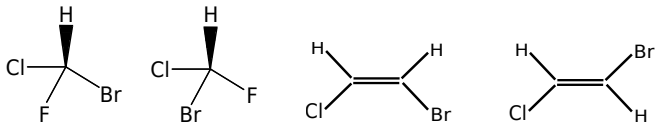
†GREYC UMR CNRS 6072, ‡LCMT UMR CNRS 6507,
Caen, France
{pierre-anthony.grenier,didier.villemin}@ensicaen.fr,
luc.brun@greyc.ensicaen.fr

Abstract. Molecules being often described using a graph representation, graph kernels provide an interesting framework which allows to combine machine learning and graph theory in order to predict molecule's properties. However, some of these properties are induced both by relationships between the atoms of a molecule and by constraints on the relative positioning of these atoms. Graph kernels based solely on the graph representation of a molecule do not encode this relative positioning of atoms and are consequently unable to predict accurately some molecule's properties. This paper presents a new method which incorporates spatial constraints into the graph kernel framework in order to overcome this limitation.

Keywords: Graph kernel, Chemoinformatics, Chirality.

1 Introduction

A molecular graph $G = (V, E, \mu, \nu)$ is a description of a molecule by a graph where the unlabeled graph (V, E) encodes the structure of the molecule, each vertex encoding an atom and each edge a bond between two atoms, μ associates to each vertex a label encoding the nature of the atom (carbon, oxygen, ...) and ν associates to each edge a type of bond (single, double, triple or aromatic). Several graph kernels [3, 1] based on this representation have been proposed in order to predict molecule's properties. However, some molecules may have a same molecular formula, a same molecular graph but a different relative positioning of their atoms inducing different properties. Such molecules are said to be stereoisomers. However, usual graph kernels based on the molecular graph representation are not able to capture any dissimilarity between such molecules. From a more local point of view, an atom is called stereocenter if a permutation of the positions of two atoms belonging to its neighborhood produces a different stereoisomer. In the same way, two connected atoms form a stereocenter if a permutation of the positions of two atoms belonging to the union of their neighborhoods produces a different stereoisomer. According to chemical experts, stereoisomerism is represented to 98% by the geometrical isomerism of double connection and the asymmetry of carbons. We thus focus the remaining of this paper on those case.



(a) Asymmetric carbon (b) Double bond

Fig. 1. Two types of stereocenters.

In order to get an intuition of stereoisomerism, let us consider an acyclic molecular graph rooted on an atom of carbon with four neighbors, each neighbor being associated to a different subtree. Such an atom, called an asymmetric carbon, is a stereocenter and has two different spatial configurations of its neighbors encoded by a same molecular graph (Fig. 1(a)). Using molecule represented in Fig. 1(a), one configuration corresponds to the case where the three atoms (Cl,Br,F) considered from the atom H are encountered in this order when turning clockwise around the central carbon atom. The alternative stereoisomer corresponds to the case where this sequence of atoms is encountered counterclockwise when considered from the same position. Two carbons, connected by a double bond, can also define stereoisomers (Fig 1(b)). Indeed, on the left side of Fig.1(b) both hydrogen atoms are located on the same side of the double bond while they are located on opposite sides on the stereoisomer represented on the right. In this case both carbon atoms of the double bond correspond to a stereocenter.

Method described in [2] includes information about the spatial configuration of atoms within the tree-pattern kernel [3]. However, this method only considers the direct neighbors of a stereocenter while, as shown by Fig. 2, the difference between two subtrees of a stereocenter may not be located on the root of the subtrees. In this last case [2] considers as identical two different stereocenters.

In this paper we propose a method to incorporate the spatial configuration of atoms within a graph kernel based on a subtree enumeration [1]. This method remains valid even when the spatial configuration is not encoded in the direct neighborhood of a stereocenter. In Section 2, we define a graph encoding of stereoisomers and we introduce chiral vertices as vertices encoding stereocenters. Next, in Section 3, we restrict our attention to acyclic molecules. Such a restriction allows us to efficiently characterise a chiral vertex by a rooted tree. In Section 4, we define the smallest tree characterizing a chiral vertex and use

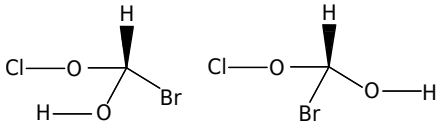


Fig. 2. Asymmetric carbons with identical neighborhood.

this information to design a graph kernel between chiral molecules. Finally, we demonstrate the validity of our kernel through experiments in Section 5.

2 Encoding of stereoisomers

An usual method in chemistry to encode stereoisometry consists in considering a fixed order on the neighborhood of each vertex. In order to encode such an information, we introduce the notion of ordered graph. An ordered graph $G = (V, E, \mu, \nu, ord)$ is a molecular graph $\hat{G} = (V, E, \mu, \nu)$ together with a function ord which maps each vertex to an ordered list of its neighbors:

$$ord \begin{cases} V \rightarrow & V^* \\ v \rightarrow & v_1 \dots v_n \end{cases} \tag{1}$$

where $V(v) = \{v_1, \dots, v_n\}$ denotes the neighborhood of v.

Two ordered graphs $G_1 = (V_1, E_1, \mu_1, \nu_1, ord_1)$ and $G_2 = (V_2, E_2, \mu_2, \nu_2, ord_2)$ are said to be isomorphic $G_1 \underset{o}{\simeq} G_2$ iff there is an isomorphism between both graphs which respects the order on the neighborhoods:

$$\exists f \in \mathrm{Isom}(\hat{G_1}, \hat{G_2}) \; s.t. \; \forall v \in V_1 \; ord_1(v) = v_1 \dots v_n, \; ord_2(f(v)) = f(v_1) \dots f(v_n) \tag{2}$$

Note that, the ordered graph isomorphism induces an equivalence relationship as well as the usual graph isomorphism.

For exemple, in Fig. 1(a) the ordered list H,Cl,Br,F for the central carbon represents the molecule to the left (and H,Cl,F,Br represents its stereoisomer). But if we consider the molecule from the Cl atom, the list Cl,H,F,Br is a valid alternative encoding of the molecule. So, a spatial configuration of atoms within a neighborhood must be encoded by several equivalent orders. We thus introduce the notion of partially ordered graph which encodes all equivalent orderings of an ordered graph. A partially ordered graph (G, Σ) is an ordered graph G with a set of re-ordering functions Σ where $\sigma \in \Sigma$ associates to each vertex v a permutation on $\{1, \dots, |V(v)|\}$. Let $G = (V, E, \mu, \nu, ord)$ be an ordered graph, $\sigma(G) = (V, E, \mu, \nu, ord_\sigma)$ is defined as the application of σ on each ordered neighborhood of G:

$$\forall v \in V \; s.t. \; ord(v) = v_1, \dots, v_n, \; ord_\sigma(v) = v_{\varphi(1)}, \dots, v_{\varphi(n)} \text{ with } \varphi = \sigma(v). \tag{3}$$

Two partially ordered graphs (G_1, Σ_1) and (G_2, Σ_2) are said to be isomorphic iff:

$$G_1 \underset{po}{\simeq} G_2 \Leftrightarrow \begin{cases} \forall \sigma_1 \in \Sigma_1, \exists \sigma_2 \in \Sigma_2 \,|\, \sigma_1(G_1) \underset{o}{\simeq} \sigma_2(G_2) \\ \forall \sigma_2 \in \Sigma_2, \exists \sigma_1 \in \Sigma_1 \,|\, \sigma_1(G_1) \underset{o}{\simeq} \sigma_2(G_2) \end{cases} \tag{4}$$

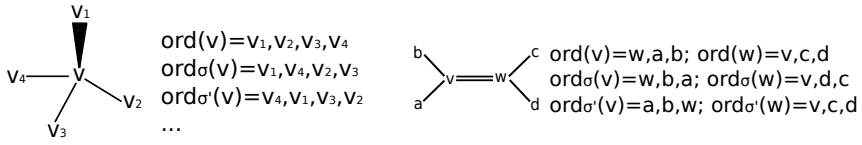
The relationship induced by partially ordered isomorphisms is reflexive and transitive as the one induced by ordered graph isomorphisms. This relation is also symmetric since we consider both re-ordering functions of Σ_1 and Σ_2. We denote by $\mathrm{IsomOrderP}(G_1, G_2)$ the set of isomorphism between two partially ordered graph G_1 and G_2.

2.1 Partially ordered graph encoding of a molecule

The partial ordered graph of a molecule is defined by first defining its molecular graph $G_{unordered} = (V, E, \mu, \nu)$. Let us denote V_{C_1} the subset of V containing all atoms of carbon with four neighbors: $V_{C_1} = \{v \in V \mid \mu(v) = \text{'}C\text{'} \text{ and } |V(v)| = 4\}$. The subset of V containing all atoms of carbon which share a double bond with another carbon is noted V_{C_2}: $V_{C_2} = \{v \in V \mid \exists e(v, w) \in E, \nu(e) = 2, |V(v)| = |V(w)| = 3 \text{ and } \mu(v) = \mu(w) = \text{'}C\text{'}\}$. For each $v \in V_{C_2}$ we denote $w = n_=(v)$ the other carbon of its double bond. In order to encode spatial configurations, let us define an ordered graph $G_{ordered}$ from $G_{unordered}$. Each vertex $v \in V - V_{C_1} - V_{C_2}$ does not require any encoding of the configuration of its neighborhood. The ordered list of its neighbors is thus set randomly. In order to set an order on the neighborhood of a vertex $v \in V_{C_1}$ we set randomly one of its neighbor v_1 at the first position. The three other neighbors of v are ordered in a way such that if we look at v from v_1, the three remaining neighbors are ordered clockwise (Section 1). One of the three orders (defined up to circular permutations) fulfilling this condition is chosen randomly (Fig. 3(a)). Finally, let us consider a vertex $v \in V_{C_2}$, with $n_=(v) = w$, $V(v) = \{w, a, b\}$ and $V(w) = \{v, c, d\}$. The order on neighborhoods of v and w are set as $ord(v) = w, a, b$ and $ord(w) = v, c, d$, whereby a, b, c, d are traversed clockwise when turning around the double bond for a given plane embedding. We add to this graph the set of re-ordering function Σ containing all the re-ordering functions σ such that: for each v in V_{C_1}, $\sigma(v)$ corresponds to an even number of transpositions on $\{1, \ldots, |V(v)|\}$ and for each v in V_{C_2}, with $n_=(v) = w$, $\sigma(v)$ and $\sigma(w)$ correspond to a same number of transpositions (Fig. 3(b)). Indeed, an additional transposition on one of the atoms of a double bond, would correspond to a permutation of the relative positioning of its neighbors hence enconding a different stereoisomer (Section 1).

Remark 1. Using the above construction scheme, the re-ordering functions of any partially order graph encoding a molecule satisfies the following properties:

- Given a sequence of neighbors of each vertex, we can always find a re-ordering such that the ordered list of each vertex starts by its selected neighbor.
- For any re-ordering functions, the permutations associated to two adjacent carbons belonging to V_{C_2} may be decomposed into a same number of transpositions.



(a) Element of V_{C_1} (b) Two elements of V_{C_2}

Fig. 3. Example of elements of V_{C_1} and of V_{C_2} with their ordered list (top) and the ordered lists obtained using two permutations $\sigma \in \Sigma$ and $\sigma' \in \Sigma$

A partially ordered graph encodes the spatial configuration of atoms within the neighborhood of each of its vertex. Let us now define a stereocenter (also called a chiral vertex).

Definition 1. Let $G = (V, E, \mu, \nu, ord, \Sigma)$ a partially ordered graph. A vertex $v \in V$ of degree n is a chiral vertex iff:

$$\forall (i,j) \in \{1,\dots,n\}^2,\ \nexists f \in \text{IsomOrderP}(G, \tau_{i,j}(G)) \text{ with } f(v) = v$$

where $\tau_{i,j}$ is a re-ordering function equals to the identity on all vertices except v for which it permutes the vertices of index i and j in $ord(v)$.

In other words, a vertex is chiral if any permutation of its neighbors produces a different partially ordered graph (called a different stereoisomer within the chemistry framework).

3 Isomorphism between labeled partially ordered tree

Let us now restrict our attention to acyclic graphs in order to obtain a more efficient calculus of isomorphisms between partially ordered graphs. Given a rooted tree, the father of each node v is denoted p_v. We define an ordered rooted tree $T = (V, E, \mu, \nu, ord)$ as a rooted tree $\hat{T} = (V, E, \mu, \nu)$ with a function ord mapping each vertex to an ordered list of its children. Like the isomorphism between ordered graph presented in Sec. 2, there is an isomorphism between two labeled ordered tree $T_1 = (V_1, E_1, \mu_1, \nu_1, ord_1)$ and $T_2 = (V_2, E_2, \mu_2, \nu_2, ord_2)$ if there is an isomorphism between both trees which complies with their order :

$$\exists f \in \text{Isom}(\hat{T_1}, \hat{T_2})\ s.t.\ \forall v \in V_1\ ord_1(v) = v_1 \dots v_n,\ ord_2(f(v)) = f(v_1) \dots f(v_n) \tag{5}$$

where $\{v_1, \dots, v_n\}$ denotes the children of v. Note that, an isomorphism between ordered tree maps roots of each tree one on the other and preserves father-child relationships.

Following [4], we associate to each ordered rooted tree, a unique depth-first string. This string is based on the sequence of node and edge labels obtained by traversing the tree in a depth-first order. As shown by [4](Lemma 2.2), two isomorphic ordered trees have the same depth-first string encoding and conversely.

Using the same approach than for partially ordered graphs, a partially ordered rooted tree (T, Σ) is an ordered rooted tree T associated to a set of re-ordering functions Σ on the children of each vertex. To define a partially ordered tree (T, Σ_T), from an acyclic partially ordered graph (G, Σ_G) encoding a molecule, we have to define a root and for each vertex an order and a set of permutations on its children encoding equivalent orders. Since the root has no parent, its children correspond to its set of neighbors and we set $ord_G(r) = ord_T(r)$. For any other vertex v, the list of its children is the list of its neighbors minus p_v. To define an order for each v in T, we apply one of the re-ordering function $\sigma \in \Sigma_G$ which puts p_v in the first position (Remark 1). The set of re-ordering

functions Σ_T is defined by considering all re-ordering functions $\sigma \in \Sigma_G$ which, for each $v \in V$, keep p_v in the first position of the ordered list of v.

In order to define a unique code for each partially ordered tree we define, as in [4], the depth-first canonical form (DFCF*) of a partially ordered tree, as the ordered tree that gives the minimal depth-first string encoding among all possible ordered trees $\sigma(T)$ obtained by applying $\sigma \in \Sigma$ on T. The depth-first string encoding of the DFCF* is called the depth-first canonical string (DFCS*) of a partially ordered tree. Since, two isomorphic ordered trees have the same depth-first string encoding, two partially ordered trees are isomorphic if their DFCS* are identical.

Given a unique code associated to a partially ordered rooted tree, the chirality of a vertex may be efficiently tested if one can transpose definition 1 to partially ordered rooted trees:

Proposition 1. Let $T = (V, E, \mu, \nu, ord, \Sigma)$ be a partially ordered tree rooted in r. r is a chiral vertex if $\forall (i,j) \in \{1, \ldots, |V(r)|\}^2$, $T \underset{po}{\not\simeq} \tau_{i,j}(T)$,

where $\tau_{i,j}$ is a re-ordering function equals to the identity on any vertex but r where it permutes children of index i and j in the ordered list of r.

Proof. Using acyclic graphs, an isomorphism between partially ordered rooted trees corresponds to an isomorphism between partially ordered graphs with an additional constraint on the mapping of both roots. If we can find an isomorphism between T and $\tau_{i,j}(T)$ such an isomorphism f satisfies $f(r) = r$ and also corresponds to an isomorphism between partially ordered graphs. Conditions of Def. 1 are thus violated and r is not chiral. The reverse implication may be demonstrated using the same type of reasoning. □

A partially ordered tree (T, Σ) can have two isomorphic subtrees whose roots have the same parent. In that case a permutation exchanging those subtrees on the DFCF* leads to an isomorphic ordered tree. In such a case, the root of these subtrees are said to be equivalent:

$$v_i \sim v_j \Leftrightarrow \begin{cases} \exists v \in V \; s.t. \; p_{v_i} = p_{v_j} = v \text{ and} \\ \exists \sigma \in \Sigma \mid \varphi(i) = j, \; \text{DFCF*}(\sigma(T)) \underset{o}{\simeq} \text{DFCF*}(T) \text{ with } \varphi = \sigma(v) \end{cases} \tag{6}$$

Since all equivalent nodes are the children of a same parent, the representative of each class is defined as the vertex with the minimal index within the ordered list of children of its parent:

$$\forall i \in \{1, \ldots, n\} \; rep(v_i) = min\{j \mid v_j \sim v_i\}. \tag{7}$$

4 From a global to a local characterization of chirality

Proposition 1 provides a global characterization of chirality. However, such a proposition does not allow to characterize the minimal subgraph of a molecule which induces the chiral property of a vertex. Using acyclic graphs, such a minimal subgraph corresponds to the smallest partially rooted tree, rooted on a chiral vertex v which allows to characterize the chirality of v using Proposition 1.

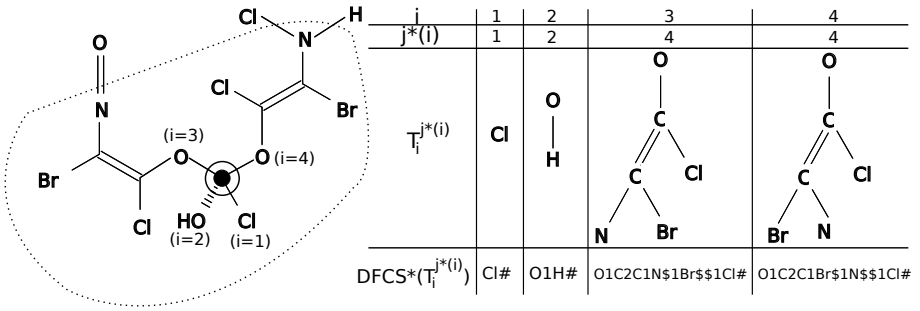


Fig. 4. Left: An asymmetric carbon ⊙with its minimal chiral subtree (*surrounded by a dotted line*). Right: minimal subtrees rooted on its children.

4.1 Minimal chiral subtree of an asymmetric carbon

Let v be a chiral vertex representing an asymmetric carbon. We denote its neighbors $v_1, \ldots, v_4$. We consider the partially ordered tree (T, Σ) rooted in v and described in Sec. 3. We note $T_1, \ldots, T_4$ the subtrees of T rooted on the children of v. For any $i \in \{1,2,3,4\}$ we denote T_i^j the subtree of T_i composed of all nodes with a depth lower than j. According to Proposition 1, the chirality of v may be characterized from its subtrees T_i^j iff all pairs of subtrees are non isomorphic. Indeed, in such a case no transposition of two subtrees T_i^j and $T_k^{j'}$ can induce an isomorphic partially ordered rooted tree. Therefore for each $i \in \{1,2,3,4\}$, we define the minimal subtree associated to v_i as $T_i^{j^*(i)}$ with $j^*(i) = \min\{j \mid \forall k \in \{1,\ldots,4\} - \{i\},\ T_i^j \not\simeq T_k^j\}$. For exemple in Fig. 4, the root of T_1 is a Cl atom and the root of each other T_i is an oxygen atom, thus $j^*(1) = 1$. The minimal chiral subtree of v is the subtree of T rooted on v, where v has for children $T_1^{j^*(1)}, \ldots, T_4^{j^*(4)}$. The asymmetric carbon is then represented by the DFCS* of this tree.

To find $j^*(i)$, we increase j for each T_i^j until $T_i^j \not\simeq T_k^j$ for each $k \in \{1,\ldots,4\}$, $k \neq i$. At each iteration we compute the DFCS* of each tree. Therefore the calculus of the minimal chiral subtree of v is performed in $\mathcal{O}((\max_i |T_i^{j^*(i)}|)^2)$ which is bounded by $\mathcal{O}(|V|^2)$.

4.2 Minimal chiral subtree of double bond

Let us consider a double bond $e = (v_a, v_b)$ and let us denote by v_a^1 and v_a^2 the two remaining neighbors of v_a. Considering the partially ordered tree T rooted on v_a, v_a is chiral only if the subtrees rooted on the children of v_a are not isomorphic (Proposition 1). This implies that the two subtrees rooted on v_a^1 and v_a^2 are not isomorphic. This necessary condition is however not sufficient. Indeed if the subtrees rooted on the remaining neighbors v_b^1 and v_b^2 of v_b are isomorphic, then one can apply a re-ordering function $\sigma \in \Sigma$ on T which simultaneously permutes the subtrees rooted on v_a^1 and v_a^2 and the subtrees rooted on v_b^1 and

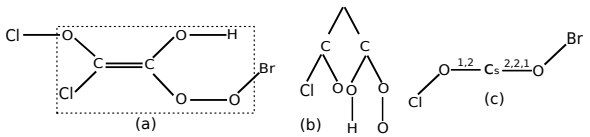


Fig. 5. A double bond (a), its minimal chiral subtree (b) and its contraction (c).

v_b^2 (Remark 1). The resulting rooted tree $\sigma(T)$ is isomorphic to T by definition but also to $\tau(T)$, where τ permutes only vertices v_a^1 and v_a^2 in the ordered list of children of v_a. In such a case, v_a is non chiral (Proposition 1). Therefore, the non chirality of v_b induces the non chirality of v_a and conversely.

Hence v_a and v_b are chirals, only if the two following conditions are satisfied: subtrees rooted on v_a^1 and v_a^2 are non isomorphic and subtrees rooted on v_b^1 and v_b^2 are also non isomorphic.

In order to encode this constraints, we define as in Section 4.1 the minimal non isomorphic subtrees rooted on v_a^1 (T_a^1) and v_a^2 (T_a^2) together with the minimal non isomorphic subtrees rooted on v_b^1 (T_b^1) and $v_b^2(T_b^2)$. We denote by T_a and T_b the two partially ordered rooted trees rooted on v_a and v_b. The subtrees of these two roots being respectively (T_a^1, T_a^2) and (T_b^1, T_b^2).

The tree encoding the chirality of the double bond is then defined as a partially ordered rooted tree, whose root corresponds to a virtual vertex (not corresponding to any atom) connected to the two subtrees T_a and T_b. As for Sec. 4.1, the computation of the minimal chiral subtree is bounded by $\mathcal{O}(|V|^2)$. Fig. 5a represents a double bond between two carbon atoms with its minimal chiral subtree (Fig. 5b).

4.3 Graph Contraction

Using results in Section 4.1 and 4.2, each stereocenter may be associated to a minimal chiral subtree and a DFCS* code representing it (Section 3). However, properties of a molecule are both determined by its set of minimal chiral subtrees and by relationships between these trees and the remaining part of the molecule. In order to obtain a local characterization of such relationships, we propose to contract the minimal chiral subtree of each stereocenters.

Let us consider a stereocenter s and its minimal chiral subtree $(T = (V_T, E_T), \Sigma)$ associated to a DFCS* code c_s. We define for this tree a set of connection vertices $V_{\text{con}} = \{v \in \text{Leaf}(T) \mid d(v) > 1\}$ and a set of edges to contract $K_T = E_T - E_{\text{con}}$ where $E_{\text{con}} = \{(v, p_v) \in V_{\text{con}} \times V_T\}$. The contraction of K_T creates a new graph $G_s = (V_s, E_s)$, with a contracted node n_s labeled by c_s and $V_s = V - (V_T - V_{\text{con}}) \cup \{n_s\}; E_s = E - K_T$ (Fig. 5c).

Each edge of E_{con} connects an element l of V_{con} to n_s in G_s. The label of $e = (n_s, l)$ has to encode the position of l in the minimal chiral subtree. We thus consider the path connecting r to l in the minimal chiral subtree: $CP(l) = v_1, .., v_n$ where $v_1 = r$ and $v_n = l$. Let us denote i_j the index of v_j in the ordered list of children of p_{v_j}. The sequence $i_2 \dots i_n$ defines a unique path

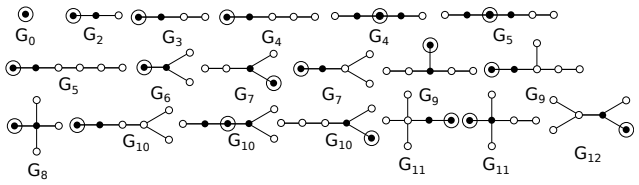


Fig. 6. The set of stereotreelet with $n_s(\odot)$, elements of $V_{con}(\bullet)$, elements of $V - V_{con}(\circ)$

in the chiral subtree associated to n_s. Such a sequence may thus be considered as a proper label of edge e. However as mentioned in Section 3, some paths may pass through equivalent subtrees and should thus be considered as equivalent. In order to encode such equivalence relationship we define the label of e as $\nu(e) = \bigodot_{i=2}^{n} rep(v_i)$ with rep defined in Eq. 7.

4.4 StereoTreelet

For each stereocenter s we have a graph G_s. The stereotreelets of G_s are defined as all subtrees of G_s whose size is lower than 6 and which include n_s. Since each neighbors v of n_s corresponds to a leaf of the minimal chiral tree of s, the edge (v, n_s) is already encoded within the code c_s of n_s. Consequently, we impose that each neighbor v of n_s in a stereotreelet must have at least another neighbor (different of n_s). This constraint induces the set of stereotreelets represented in Fig. 6. The set of stereotreelet $\mathcal{T}(G)$ of G is defined as the union of stereotreelets of each G_s.

When all stereotreelets of G have been enumerated, we compute its spectrum $s(G)$ which corresponds to a vector representing the treelet distribution. Each component of this vector is equal to the frequency of a given stereotreelet t: $s(G) = (f_t(G))_{t \in \mathcal{T}(G)}$ with $f_t(G) = |(t \subseteq G)|$. The kernel between two graphs G and G' is defined as a sum of kernels between the different number of treelets common to both graphs: $k(G, G') = \sum_{t \in \mathcal{T}(G) \cap \mathcal{T}(G')} K(f_t(G), f_t(G'))$.

5 Experiments

We have tested our method on a dataset of acyclic chiral molecules [5] related to a regression problem. This dataset is composed of 90 molecules together with their optical rotations. In practice, we only select 35 molecules, since almost all molecules have only one stereocenter, and for 55 molecules this stereocenter is unique in the dataset. Such molecules correspond to a property represented only once in the dataset which can thus not be accurately predicted. The property to predict, the optical rotation, is connected with chirality and has a standard deviation of 38.25 for the 35 selected molecules.

For our experiment we use a leave-one-out cross-validation on the dataset to predict the optical rotation of each molecule. The predicted rotations are

Table 1. Optical rotation prediction for the acyclic chiral dataset.

Method	Kernel Ridge		Weighted Average		Gram's matrix
	Average Error	RMSE	Average Error	RMSE	computations (s)
Random Kernel	31.7	39.5	32.0	39.3	0.03
KMean[6]	31.0	38.7	32.3	39.6	153.84
Treelet Kernel[1]	26.0	33.9	28.9	37.4	0.49
Stereotreelet Kernel	21.0	25.6	11.6	16.3	0.13

computed by using both kernel ridge regression and the weighted mean of known values using the similarity measure provided by the kernel $\hat{y} = \frac{\sum_i k(G_i,G)\times y_i}{\sum_i k(G_i,G)}$. We present in Table 1 the average errors, Root Mean Squared Errors (RMSE) and computation times of the Gram matrix for our method and the ones of [6, 1] which do not take into account chirality. Results obtained by using a random Gram matrix are also shown.

Weighted mean provides much better results for our kernel since on this dataset each molecule has a non null similarity with a reduced number of molecules (less than 10). Such a reduced number of data do not allow kernel ridge regression to perform reliable prediction. Other methods provide similar results than those obtained using a random Gram matrix. These results are also comparable with the variance of the dataset. Such a result may be explained by the fact that optical rotation is connected to chirality which is not encoded by these kernels.

6 Conclusion

In this paper we proposed a graph kernel for chemoinformatics that considers the spatial constraints of atoms within molecules. Our experiments show promising results and our future work will consist to create larger datasets and to extend our method to graphs including cycles.

References

1. Gaüzère, B., Brun, L., and Villemin, D. *Two new graphs kernels in chemoinformatics.* In Pattern Recognition Letters, April 2012.
2. Brown, J., Urata, T., Tamura, T., Arai, M., Kawabata, T., and Akutsu, T. *Compound analysis via graph kernels incorporating chirality* J. Bioinform. Comp. Bio. S1: 63-81, 2010.
3. Mahé, P., and Vert, J.-P. *Graph kernels based on tree patterns for molecules.* Machine Learning, 75(1):3-35, October 2008.
4. Chi, Y., Yang, Y., and Muntz, R. R. *Canonical forms for labeled trees and their applications in frequent subtree mining.* KAIS, 8(2):203-234, 2005.
5. Zhu, H. J., Ren, J., and Pittman C. U., Jr. *Matrix model to predict specific optical rotations of acyclic chiral molecules.* Tetrahedron 2007, 63, 2292-2314.
6. Suard, F., Rakotomamonjy, A., and Bensrhair., A. *Kernel on bag of paths for measuring similarity of shapes.* In ESANN, 2002.