Candidature concours CR CNRS 2021
Jules Vidal

11 janvier 2022

Table des matières

1 Contexte

La taille et la complexité toujours croissantes des ensembles de données produites en sciences constituent un défi majeur pour leur interprétation. Pour résoudre ces problèmes, de nouvelles techniques sont développées en analyse de données pour arriver à capturer efficacement les caractéristiques principales de grands ensembles de données, afin de permettre leur visualisation et leur analyse interactive. Les outils développés dans le domaine de l'analyse topologique de données servent précisément à cela. Ils forment une famille de techniques axés sur l'extraction générique, robuste et efficace de caractéristiques structurelles dans les données [22]. Au cours des dernières années, de nombreuses méthodes d'analyse et de visualisation de données ont été construites autour de ces concepts [33], et appliquées à un large spectre de domaines, de l'astrophysique [59, 57], à la biologie et l'imagerie médicale [8, 3] en passant par la chimie [5, 26, 47], la mécanique des fluides [38], les sciences des matériaux [31, 29, 58], ou la combustion turbulente [41, 10, 28]. Dans le cas des données scalaires, les algorithmes d'analyse topologique de données reposent sur des **signatures topologiques** établies, telles que des arbres de contour [9, 18, 60, 11], les graphes de Reeb [52, 6, 48] ou les complexes de Morse-Smale [19, 27, 55, 30]. Des exemples de signatures topologiques calculées sur des champs scalaires sont donnés en figure 1. En particulier, le diagramme de persistance [23] est une représentation concise des données, qui résume visuellement la répartition de caractéristiques d'intérêt dans un jeu de données, en fonction d'une mesure d'importance appelée **persistance topologique**. Sa concision, combinée à sa stabilité, a popularisé son utilisation en apprentissage automatique [12, 53, 54] et en analyse de données interactive, puisqu'il fournit rapidement des informations sur le nombre et l'importance de traits caractéristiques dans les données, comme illustré en figure 2.

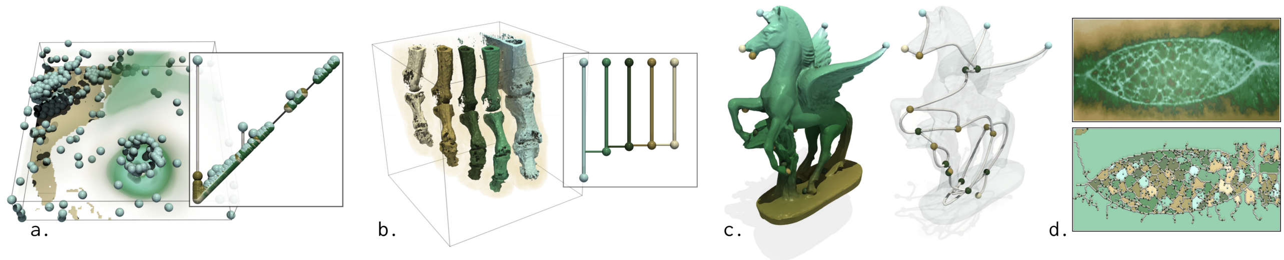

FIGURE 1 – Exemples de signatures topologiques calculées sur des données scalaires. **a.** : diagramme de persistance calculé sur le résultat d'une simulation d'ouragan. **b.** : un arbre de jointure calculé sur une image médicale (le scan d'un pied). **c.** : le graphe de Reeb d'un champ harmonique sur une surface fermée. **d.** : le complexe de Morse-Smale d'une image acquise par microscopie électronique. Exemples et données tirés de [62].

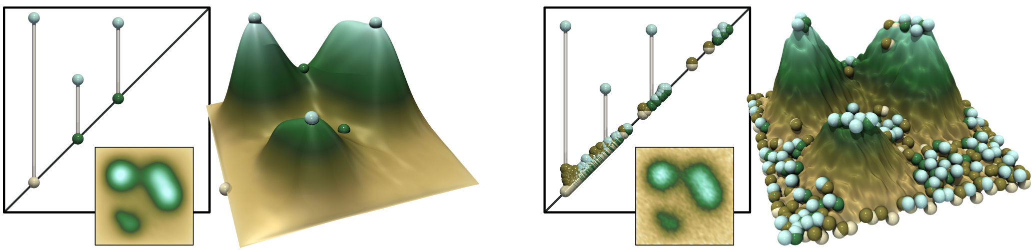

FIGURE 2 – Diagrammes de persistance d'un champ scalaire lisse (à gauche) et bruité (à droite). Les sphères indiquent la localisation des points critiques : en cyan les maxima, en beige les minima, et en brun et vert les points selles. Les trois montagnes principales sont clairement visibles dans les diagrammes (grandes paires de persistance), tandis que les petites paires près de la diagonale indiquent des structures correspondant à du bruit dans les données. La hauteur de chaque paire indique la **persistance topologique** de la structure associée.

2 Rapport d'activités

2.1 Recherche

Cette section présente les travaux de recherche effectués pendant mon doctorat, qui s'est déroulée de septembre 2018 à décembre 2021. Ma thèse, intitulée "Progressivité en Analyse Topologique de données", visait à développer des méthodes progressives pour l'analyse topologique de données, c'est-à-dire des méthodes qui fournissent une approximation exploitable du résultat qui sera raffinée par itérations successives. Dans une configuration interactive, les algorithmes progressifs peuvent améliorer l'expérience utilisateur de deux manières. Premièrement, les utilisateurs peuvent laisser ces algorithmes affiner progressivement leurs résultats, tout en recevant des retours visuels en continu, et les arrêter lorsque les sorties sont jugées satisfaisantes. Deuxièmement, les utilisateurs peuvent également définir a priori une limite supérieure sur le temps de calcul, après quoi le calcul est interrompu. Cela permet de concevoir des systèmes interactifs avec un temps de réponse garanti. Les algorithmes progressifs peuvent également être utiles pour les configurations non interactives, telles que les calculs en mode batch sur des ordinateurs hautes performances, où l'allocation des ressources informatiques doit souvent être finement contrôlée. Dans ce contexte, la progressivité permet d'affecter un budget de temps de calcul précis aux programmes d'analyse de données. Ceci est particulièrement pertinent pour les applications où le temps de calcul est critique telles que la simulation numérique pour la prise de décision urgente, qui était le contexte de ma thèse (voir sous-section 2.3).

2.1.1 Barycentres de Wasserstein progressifs de diagrammes de persistance

Dans ce travail, nous avons introduit un algorithme efficace pour l'approximation progressive de barycentres de Wasserstein de diagrammes de persistance, avec des applications à l'analyse visuelle de données d'ensemble. Étant donné un ensemble de champs scalaires, notre approche permet le calcul d'un diagramme de persistance qui est représentatif de l'ensemble, et qui transmet visuellement le nombre, les plages de données et les saillances des principales caractéristiques d'intérêt trouvées dans l'ensemble. De tels diagrammes représentatifs sont obtenus en calculant explicitement le barycentre de Wasserstein discret de l'ensemble des diagrammes de persistance, une tâche notoirement coûteuse en calcul. Plus précisément, nous avons revisité des algorithmes efficaces pour l'approximation de distance de Wasserstein [39, 4] pour étendre des travaux antérieurs sur l'estimation de barycentres [63]. Nous avons présenté un nouvel algorithme rapide, qui fournit un résultat se rapprochant progressivement d'un barycenter en augmentant itérativement la précision du calcul ainsi que le nombre de structures persistantes dans le diagramme de sortie. Une telle progressivité améliore considérablement la convergence et permet de concevoir un algorithme interruptible, capable de respecter des contraintes de temps de calcul. Cela permet l'approximation des barycentres de Wasserstein dans des temps interactifs.

Nous avons étendu cette méthode à technique de clustering d'ensembles où nous revisitons l'algorithme du *k-means* pour exploiter nos barycentres et calculer, dans les limites du temps d'exécution, des clusters significatifs de données d'ensemble avec leur barycentre. Des expériences effectuées sur des ensembles de données synthétiques et réelles ont montré que notre algorithme converge vers des barycentres qui sont qualitativement exploitables par rapport aux applications, et quantitativement comparable aux résultats de techniques de référence, tout en offrant une accélération d'un ordre de grandeur, même sans interruption du calcul.

Les contributions de l'approche sont illustrées en figure 3. Ce travail a fait l'objet d'une publication qui a été récompensée à la conférence IEEE VIS (Best paper honorable mention award).Le code et les données associés sont disponibles sur github : github.com/julesvidal/wasserstein-pd-barycenter

Publication associée :
"Progressive Wasserstein Barycenters of Persistence Diagrams", **Jules Vidal**, Joseph Budin, Julien Tierny, *IEEE Transaction on Visualization and Computer Graphics, Proc. of IEEE VIS 2019* (Best Paper Honorable Mention Award) [65]

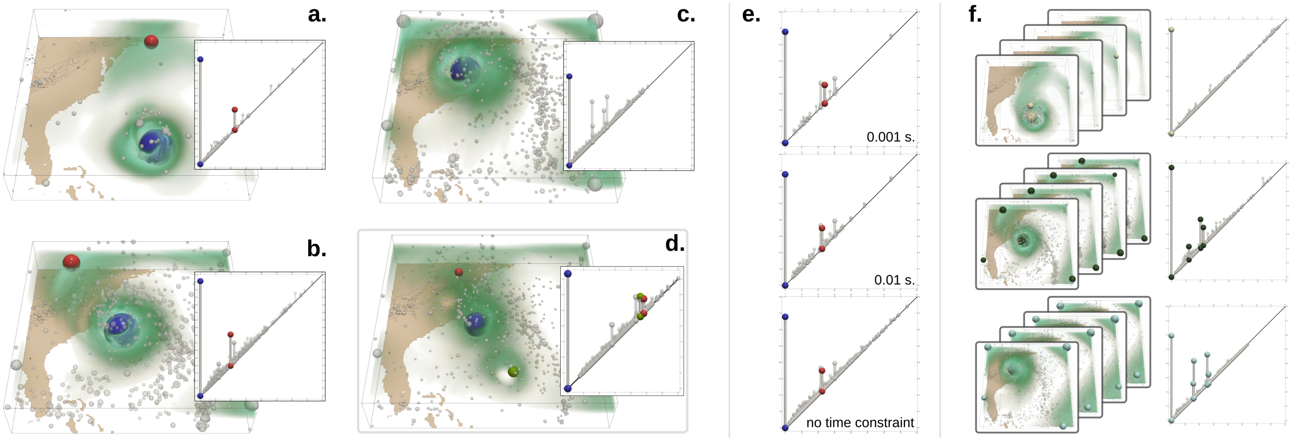

FIGURE 3 – Les diagrammes de persistance de trois membres (a-c) de l'ensemble Isabel (calculés sur la norme de la vitesse du vent) encodent de manière concise et visuelle le nombre, la plage de valeurs et l'importance des caractéristiques d'intérêt trouvées dans les données (respectivement le mur de l'oeil de l'ouragan et une autre région de grand vent, représenté par des sphères bleu et rouge dans (a)). Dans ces diagrammes, les structures avec une persistance inférieure à 10% de la plage de fonction ou du bord du domaine sont affichées en blanc transparent. La moyenne point-à-point pour ces trois membres (d) présente trois caractéristiques intérieures importantes (dues à des emplacements distincts du mur de l'oeil du cyclone, aux endroits des sphères bleue, verte et rouge), bien que les diagrammes des membres d'entrée ne rapportent que deux structures saillantes au plus, situées à différentes plages de données (la structure rouge est plus bas sur la diagonale en (a) et (b)). Le barycentre de Wasserstein de ces trois diagrammes (e) fournit une vue plus représentative des caractéristiques trouvées dans cet ensemble, car il rapporte un nombre de structure, une plage de valeurs et une persistance qui correspond mieux aux diagrammes d'entrée (a-c). Notre avons introduit un nouvel algorithme d'approximation progressive de tels barycentres, avec des temps de convergence praticables. Notre technique rend possible la spécification de contraintes de temps de calcul (e) qui permettent l'approximation de barycentres de Wasserstein en des temps interactifs. Nous présentons une application au clustering des membres de l'ensemble en fonction de leur diagramme de persistance (f), ce qui permet l'exploration visuelle des principales tendances des structures topologiques trouvées dans l'ensemble.

2.1.2 Une approche progressive à l'analyse topologique de champs scalaires

Nous avons introduit dans ces travaux les premiers algorithmes progressifs pour l'analyse topologique de données scalaires.

Notre approche est basée sur une représentation hiérarchique des données d'entrée et la notion clé de **sommets topologiquement invariants**, que j'ai définie dans le cadre de ma thèse. Il s'agit de sommets qui n'ont pas d'impact sur la description topologique des données et pour lesquels nous avons montré qu'aucun calcul n'est nécessaire lorsqu'ils sont introduits dans la hiérarchie. Cette notion permet de définir des algorithmes topologiques *coarse-to-fine* efficaces, qui exploitent des mécanismes de mise à jour rapide pour les sommets ordinaires et évitent le calcul pour les sommets topologiquement invariants. Nous illustrons notre approche avec deux exemples de algorithmes topologiques (extraction de points critiques et calcul de diagramme de persistance), qui génèrent des sorties interprétables sur interruption à la demande et qui les affinent progressivement autrement. Des expériences sur des ensembles de données réels montrent que notre stratégie, en plus du retour visuel continu qu'elle fournit, améliore même les performances d'exécution en comparaison de méthodes non-progressives. Des accélérations supplémentaires du calcul sont rendues possibles grâce à des techniques de parallélisme à mémoire partagée.

Nous avons montré l'utilité de notre approche en mode batch et dans des configurations interactives, où elle permet respectivement le contrôle du temps d'exécution de pipelines d'analyse topologique complets ainsi que des aperçus des caractéristiques topologiques trouvées dans un jeu de données, avec des mises à jour progressives fournies dans des délais interactifs.

Les contributions de notre approche sont illustrées en figure 4. Le code et les données associés sont disponibles sur github : github.com/julesvidal/progressive-scalar-topology

Publication associée :
"A Progressive Approach to Scalar Field Topology", **Jules Vidal**, Pierre Guillou, Julien Tierny, *IEEE Transaction on Visualization and Computer Graphics*, 2021 [66]

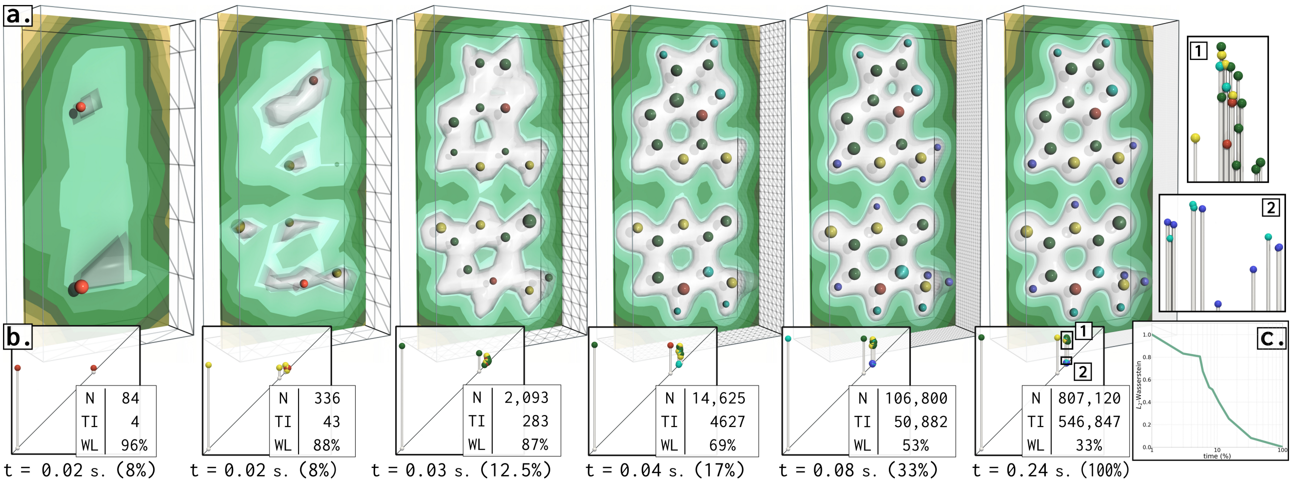

FIGURE 4 – Diagrammes de persistance progressifs (paires selle-maximum) de la densité électronique du système moléculaire adénine-thymine (AT) (une isosurface montre les deux molécules), pour quelques étapes du calcul progressif. Notre approche *coarse-to-fine* affine efficacement le diagramme de persistance en progressant le long d'une représentation hiérarchique des données d'entrée (de gauche à droite). Les maxima (désignant les atomes) sont indiqués dans le domaine (a) avec des sphères, mises à l'échelle par persistance topologique et colorées par durée de vie dans la hiérarchie des données (du rouge au bleu foncé). Dans cet exemple, le diagramme de persistance (b) capture progressivement les principales caractéristiques des données. A partir de 8% du temps de calcul (le plus à gauche), deux maxima persistants sont capturés, indiquant la présence des deux molécules. Au fur et à mesure que le calcul progresse, les atomes sont progressivement capturés, les atomes les plus lourds en premier : les atomes d'oxygène, puis d'azote et de carbone, et enfin d'hydrogène sont respectivement tous captés à partir de 12,5%, 17% et 33% du temps de calcul. À ce stade le diagramme est complet et sa précision est alors améliorée jusqu'au résultat final et exact (le plus à droite). Cette progression qualitative se confirme quantitativement par la convergence empirique de la distance L2-Wasserstein vers la sortie finale (c), qui est monotone décroissante : plus de temps de calcul donne en effet plus de précision. Nos algorithmes exploitent des mécanismes de mise à jour efficaces et des sommets topologiquement invariants (TI), qui peuvent être rapidement identifiés et pour lequel nous montrons qu'aucun calcul n'est requis, réduisant ainsi significativement la charge de travail (WL) de l'algorithme avec le temps. Globalement, notre approche progressive calcule efficacement le diagramme de persistance des données, tout en fournissant en permanence un retour visuel pertinent.

2.1.3 Approximation de diagrammes de persistance avec garanties

Ces travaux se sont déroulés durant ma dernière année de thèse. Nous avons introduit un algorithme pour l'approximation efficace du diagramme de persistance d'un champ scalaire, avec garanties, afin de pallier la principale limitation de nos travaux précédents.

Les méthodes progressives que nous avions développées précedemment [66] fournissent des descriptions approximatives de la topologie des données lorsqu'elles sont interrompues. Cependant, aucune garantie théorique ne peut être fournie quant à la qualité de cette approximation. Dans ce travail, nous avons revisité le cadre progressif présenté précédemment et introduit un nouvel algorithme d'approximation du diagramme dont l'erreur est contrôlée par l'utilisateur, spécifiquement sur la distance de *Bottleneck* [14] à la solution exacte. Notre approche est basée sur la même représentation hiérarchique des données d'entrée, et s'appuie sur des simplifications du champ scalaire pour accélérer le calcul, tout en maintenant une borne contrôlée sur l'erreur de sortie. La localité de notre approche permet une parallélisation par mémoire partagée. Des expériences menées sur des jeux de données réels montrent que pour un tolérance d'erreur légère (distance relative de Bottleneck de 5%), notre approche améliore les performances d'exécution de 18% en moyenne (et jusqu'à 48% sur des ensembles de données volumineux et très bruités) par rapport à des implémentations de référence et publiquement disponibles fournissant des résultats exacts. En plus des garanties sur son erreur d'approximation, nous montrons que notre algorithme fournit également en pratique des sorties qui sont en moyenne 5 fois plus précises (en termes de la distance L2-Wasserstein) qu'une approximation naïve de base. Nous illustrons l'utilité de notre approche pour l'exploration interactive de données et nous documentons des stratégies de visualisation pour rendre compte des incertitudes liées à notre approximation.

Les contributions de notre approche sont illustrées en figure 5. Le code et les données associés sont disponibles sur github : github.com/julesvidal/persistence-diagram-approximation

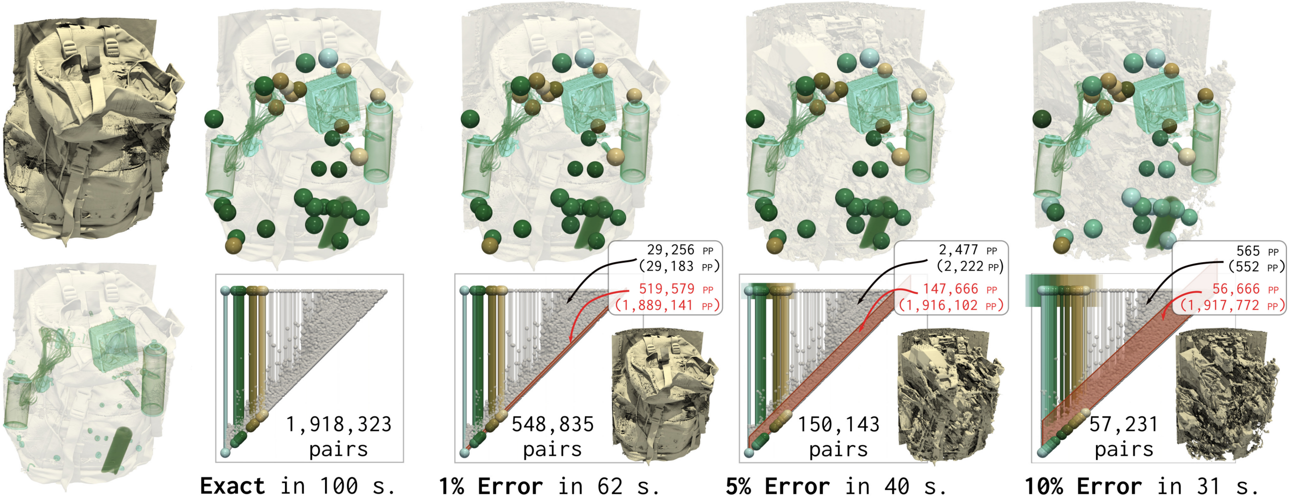

FIGURE 5 – Approximations de diagrammes de persistance pour le scan d'un sac à dos, avec différentes erreurs d'approximation. Les structures avec une grande persistance correspondent aux objets à haute densité présents dans le sac (en bas, à gauche). Dans cet exemple, notre approximation réduit le temps de calcul de 70% pour une tolérance d'erreur de 10%. Notre algorithme fournit également une approximation du champ scalaire qui reste précise autour des structures persistantes (en haut, rendu volumique des champs approchés) et détériorée ailleurs (en bas, isocontours capturant le tissu du sac). Pour chaque approximation, les trente maxima les plus persistants sont représentés par des sphères (rang du haut) qui capturent correctement les caractéristiques des données. L'incertitude résultant de notre approximation est visualisée dans le diagramme avec (i) des carrés de couleur, qui délimitent l'emplacement correct des caractéristiques *certaines* (qui sont garanties d'être présentes dans le résultat exact). Ceux-ci sont situés à l'extérieur d'une bande rouge (ii) (à proximité de la diagonale), qui indique des structures qui pourraient ne pas exister dans le résultat exact. Nos approximations capturent précisément les caractéristiques de haute persistance des données (hors de la bande rouge, nombres en noir) et collectent nettement moins de structures correspondant à du bruit (nombres en rouge, bande rouge) que le résultat exact (nombres entre parenthèses).

Publication associée :
"Fast Approximation of Persistence Diagrams with Guarantees", **Jules Vidal**, Julien Tierny, *IEEE Symposium on Large Data Analysis and Visualization*, 2021 [67]

2.2 Développement logiciel et dissémination

Mon travail de recherche est accompagné d'une attention particulière au développement et à la dissémination des technologies associées. Ainsi, tous les travaux qui ont été réalisés pendant ma thèse sont réplicables : une implémentation open-source de nos méthodes est librement disponible, ainsi que les données utilisées dans les publications. Lorsque nos travaux étaient éligibles, ils ont été certifiés réplicables dans le cadre du *Graphical Replicability Stamp Initiative* (www.replicabilitystamp.org).

Je suis également un des développeurs de **The Topology ToolKit** (TTK, topology-tool-kit.github.io)[61, 7], une librairie open-source pour l'analyse topologique et la visualisation scientifique. Bien que principalement développée à Sorbonne Université sous la direction de Julien Tierny, TTK regroupe désormais plus d'une trentaine de contributeurs issus de plus de 16 institutions académiques et industrielles à travers le monde. TTK s'intègre dans ParaView [1], un populaire logiciel de visualisation largement utilisé dans les domaines d'applications, open-source, qui est développé autour de la bibliothèque VTK (Visualization ToolKit). La bibliothèque TTK est activement développée et maintenue. Depuis 2021, elle est officiellement intégrée dans ParaView, ce qui favorise l'accessibilité et la dissémination des technologies développées dans notre communauté de recherche. Toutes les méthodes développées pendant ma thèse ont été intégrées dans TTK, et sont très facilement utilisables dans ParaView.

J'attache une grande importance à cet effort de développement et de maintenance logicielle, auquel je participe depuis le début de ma thèse. Je continuerai à intégrer mes travaux futurs dans TTK, afin qu'ils restent accessibles à la fois pour ma communauté de recherche, pour favoriser leur reproductibilité et leur extension, et à la fois pour les utilisateurs des domaines d'application.

J'ai aussi participé à l'animation de plusieurs tutoriels sur les possibilités et l'utilisation de TTK. Ces tutoriels, dont la liste est décrite en sous-section 2.4, ont été présentés à la conférence internationale IEEE VIS.

2.3 Implication dans le projet européen VESTEC

Ma thèse s'est déroulée dans le cadre du projet européen VESTEC (Visual Exploration and Sampling Toolkit for Extreme Computing, https://vestec-project.eu/). Le projet vise à développer une chaine d'outils flexible de simulation et d'analyse de données pour aider à la prise de décision urgente en cas de situation de crise. Le projet regroupe des partenaires internationaux de multiples domaines : le calcul haute performance, la simulation numérique, l'analyse et visualisation de données, le rendu, et des experts de domaines d'applications (analyse de feux de forêt, analyse de l'évolution de maladies transmises par les moustiques, et analyse de la météorologie spatiale).

Concrètement, étant donné un scénario catastrophique (par exemple un feu de forêt, un tremblement de terre, la propagation d'une épidémie), le système donne la possibilité à un utilisateur preneur de décision de lancer des ensembles de simulations numériques sur un supercalculateur. Les résultats de ces simulations sont ensuite analysés et visualisés à distance et de manière interactive afin d'évaluer les évolutions possibles de la situation et la qualité des réponses envisagées. The Topology ToolKit (sous-section 2.2) a été utilisée dans ce contexte pour effectuer une **réduction** des données d'ensemble générées par les simulations, via le calcul in-situ de diagrammes de persistance, et pour l'analyse et la visualisation topologique d'ensembles. L'urgence des cas d'applications du projet nécessite l'utilisation de méthodes d'analyse **interactives** et **in-situ** de données. Ceci a motivé le développement des techniques progressives d'analyse topologique de données présentées plus haut (sous-section 2.1).

Depuis le début de ma thèse, je travaille donc en interaction avec les différents partenaires, afin d'intégrer nos algorithmes d'analyse topologique de données dans le pipeline de simulation, d'analyse et de visualisation de données de VESTEC. Ce travail consiste à participer activement aux réunions, à présenter des comptes-rendu et à rédiger des rapports techniques. Une grande partie de ce travail s'apparente à du transfert de technologie vers les domaines d'applications. Il a fallu évaluer avec les experts des domaines applicatifs comment les méthodes topologiques pouvait profiter à leur analyse, et les aider à mettre en place les moyens techniques d'utiliser nos méthodes.

Le projet VESTEC se terminera dans le courant de l'année 2022, année pendant laquelle le prototype final du système sera présenté à de potentiels utilisateurs pour évaluation. Plusieurs publications sont en cours de préparation avec les partenaires du projet.

2.4 Liste complète de publications

Premier auteur

Journaux

— *A Progressive Approach to Scalar Field Topology*
Jules Vidal, Pierre Guillou, Julien Tierny
IEEE Transactions on Visualization and Computer Graphics, 2021
To be presented at IEEE VIS 2021

— *Progressive Wasserstein Barycenters of Persistence Diagrams*
Jules Vidal, Joseph Budin, Julien Tierny
IEEE Transactions on Visualization and Computer Graphics
Proc. of IEEE VIS 2019
Best Paper Honorable Mention Award

Conférences

— *Fast Approximations of Persistence Diagrams with Guarantees*
Jules Vidal and Julien Tierny
Proc. of IEEE Symposium on Large Data Analysis and Visualization (LDAV) 2021

Autres publications

Journaux

— *Wasserstein Distances, Geodesics and Barycenters of Merge Trees*
Mathieu Pont, **Jules Vidal**, Julie Delon, Julien Tierny
IEEE Transactions on Visualization and Computer Graphics
Proc. of IEEE VIS 2021

Conférences

— *Statistical Parameter Selection for Clustering Persistence Diagrams*
Max Kontak, **Jules Vidal**, Julien Tierny
Proc. of Super Computing workshop on Urgent HPC 2019

— *An Overview of the Topology Toolkit*
Talha Bin Masood, Joseph Budin, Martin Falk, Guillaume Favelier, Christoph Garth, Charles Gueunet, Pierre Guillou, Lutz Hofmann, Petar Hristov, Adhitya Kamakshidasan, Christopher Kappe, Pavol Klacansky, Patrick Laurin, Joshua A. Levine, Jonas Lukasczyk, Daisuke Sakurai, Maxime Soler, Peter Steneteg, Julien Tierny, Will Usher, **Jules Vidal** and Michal Wozniak
Proc. of TopoInVis 2019

Tutoriels

— *Topological Analysis of Ensemble Scalar Data with TTK*
Christoph Garth, Charles Gueunet, Pierre Guillou, Lutz Hofmann, Joshua A Levine, Jonas Lukasczyk, Julien Tierny, **Jules Vidal**, Bei Wang, Florian Wetzels
IEEE VIS Tutorials 2021

— *Topological Data Analysis Made Easy with the Topology ToolKit, What is New ?*
Martin Falk, Christoph Garth, Charles Gueunet, Pierre Guillou, Attila Gyulassy, Lutz Hofmann, Christopher Kappe, Joshua A Levine, Jonas Lukasczyk, Julien Tierny, **Jules Vidal**
IEEE VIS Tutorials 2020

— *Topological Data Analysis Made Easy with the Topology ToolKit, A Sequel*
Martin Falk, Christoph Garth, Charles Gueunet, Joshua A Levine, Jonas Lukasczyk, Julien Tierny, **Jules Vidal**
IEEE VIS Tutorials 2019

3 Projet de recherche

Mon projet de recherche consiste à développer le domaine de l'analyse topologique de données pour la visualisation, une thématique actuellement peu portée dans les UMR CNRS. Cette thématique se place à la croisée de plusieurs disciplines : la topologie computationnelle, la modélisation géométrique, et la visualisation de donnnées.

Mon projet s'articule autour de trois axes de recherche. Le premier axe, dans la continuité de mon travail de thèse, vise au développement de techniques pour le calcul interactif de représentation topologiques de données. Le second axe est concentré autour d'une problématique d'actualité dans la communauté, d'un grand intérêt pour les applications : la reconstruction de données à partir de signatures topologiques. Enfin le troisième axe vise à développer de nouvelles méthodes de visualisation d'information topologique, afin de faciliter le transfert de ces outils de visualisation vers les domaines d'applications. À plus long terme, j'aimerais m'impliquer dans le développement d'environnements interactifs pour la visualisation scientifique, et dans l'étude de l'impact environnemental de méthodes progressives et approchées en sciences des données.

Je décris ci-après les différents axes de mon projet de recherche. Cette description est suivie par mes deux vœux d'intégration :

— Dans l'équipe ORIGAMI du Laboratoire d'InfoRmatique en Image et Systèmes d'Information (LIRIS, UMR 5205)

— Dans l'équipe STORM de l'Institut de Recherche en Informatique de Toulouse (IRIT, UMR 5505)

Axe 1 : Calcul interactif d'abstractions topologiques

Motivation

a) Calcul d'abstractions topologiques Les abstractions (ou signatures) topologiques telles que le diagramme de persistance, l'arbre de jointure, le graphe de Reeb ou le complexe de Morse-Smale sont des structures de données réduites qui encodent des caractéristiques structurelles topologiques présentes à toutes les échelles dans les données. En tant que telles, elles permettent une **réduction** efficace des données initiales, et peuvent servir à des fins d'analyse ou de visualisation. Pour exemple, un diagramme de persistance est typiquement trois ordres de grandeurs plus léger que la donnée initiale. Parce qu'elles encodent des informations topologiques, ces abstractions caractérisent des structures d'intérêt dans les données indépendamment de leur orientation et de leur déformation. Elles peuvent donc être utilisées pour comparer ou classifier des données sur des critères purement topologiques. Enfin la **persistance topologique** [22] permet de donner une mesure d'importance sur les structures présentes dans les données, ce qui permet de construire des signatures topologiques stables [14] et robustes au bruit.

b) Interactivité Dans son article fondateur sur les temps de réponse des systèmes interactifs, Miller [45] a étudié l'impact du temps de réponse sur la capacité des utilisateurs à rester concentrés sur une tâche donnée. Pour des temps de réponse inférieurs à une seconde (latence préservant la continuité), le système apparaît pleinement réactif aux réglages de l'utilisateur, permettant des sessions vraiment interactives. Pour des temps de réponse inférieurs à quelques secondes (latence préservant le flux), les utilisateurs parviennent toujours à maintenir leur concentration mais les pauses dans l'exploration, en raison du calcul, mettent à mal les compétences d'interprétation de l'utilisateur. Au-dessus de dix secondes (seuil de latence préservant l'attention), les utilisateurs ont tendance à se désengager de la tâche pour poursuivre d'autres activités en parallèle, ce qui devient préjudiciable au processus d'interprétation.

Verrous

Bien que les algorithmes de base de d'analyse topologique de données aient des complexités asymptotiques raisonnables (généralement de linéaire à quadratique), la construction d'abstractions topologiques telles que le diagramme de persistance, l'arbre de jointure, le graphe de Reeb, ou le complexe de Morse-Smale peut encore nécessiter des temps de calcul importants pour des jeux de données réels. Ainsi, ces méthodes peuvent devenir un sérieux goulot d'étranglement au sein d'un système d'analyse de données interactif. Ceci devient une préoccupation dans un contexte d'exploration interactive de données, dans lequel les utilisateurs peuvent avoir besoin de plusieurs minutes en pratique pour obtenir un retour lorsqu'ils ajustent les paramètres d'une méthode d'analyse topologique. De plus, il est difficile d'estimer a priori le temps d'exécution du calcul d'une signature topologique. En effet, ce temps est hautement sensible à la taille de la sortie, c'est-à-dire à la complexité de la topologie des données, qui n'est pas connue initialement. Cette taille dépend en particulier du niveau de bruit présent dans les données.

De fait, la majorité de l'information topologique calculée décrit de petites structures dans les données, qui sont largement associées à du bruit, comme illustré en figure 6 sur un jeu de données réel. Cette fraction du résultat représente également la majeure part du temps de calcul, qui limite l'interactivité des méthodes d'analyse topologique de données.

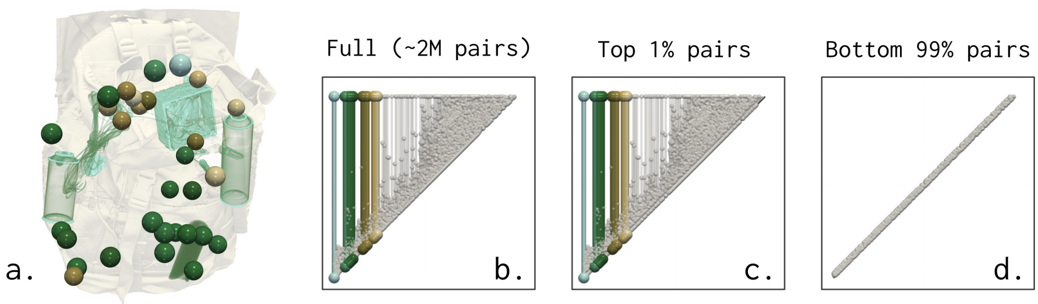

FIGURE 6 – Diagramme de persistance du champ de densité dans une acquisition de sac à dos par tomodensitométrie (a.). Le diagramme complet (b.) possède environ deux millions de paires de persistance. Le diagramme constitué du top 1% de ces paires (les 20 000 plus persistantes, c.), est presque indiscernable du résultat exact et décrit globalement la même topologie. Le restant des paires, les 99% les moins persistantes (d.) sont des très petites paires qui correspondent à du bruit dans les données, mais constituent la majeure partie du diagramme.

Directions de recherche

a) Calcul progressif Une des solutions envisageables pour le calcul interactif d'abstractions topologiques est de se tourner vers des approches progressives. Dans ce contexte, un algorithme progressif est une technique capable de s'interrompre sur demande pour fournir une sortie **approchée** exploitable, et de l'affiner autrement, convergeant progressivement vers la solution finale. La notion d'analyse de données **progressive** a été explorée par plusieurs auteurs en visualisation d'informations [68, 24, 70, 36], cependant elle a pour l'instant été peu appliquée en visualisation scientifique. Ce paradigme s'applique pourtant particulièrement bien à l'analyse topologique de données. En effet, comme expliqué plus haut, l'essentiel de l'information calculée par les méthodes topologiques concerne des petites structures d'un intérêt limité dans les données, et qui peuvent donc être approchées dans le calcul. De plus, la **persistance topologique** donne une mesure d'importance multi-échelle sur ces structures, et permet donc d'évaluer la pertinence de l'approximation calculée. Une partie de ma thèse a consisté à explorer cette piste pour le calcul du diagramme de persistance en visualisation scientifique, notamment à l'aide d'une représentation hiérarchique multi-résolution des données, avec des résultats prometteurs [66] dans le cas de champs définis sur des grilles régulières. En particulier, l'approche progressive développée dans ce travail fournit un résultat **exact** plus rapidement que des méthodes classiques, *en plus* de fournir des résultats intermédiaires exploitables de manière interactive. Cette direction de recherche originale, que j'ai moi-même initiée lors de ma thèse, mérite d'être continuée au-delà. Ces résultats préliminaires peuvent notamment être étendus à des signatures topologiques plus explicites, comme les arbres de jointure, et à des données plus génériques, en particulier des champs scalaires définis sur des triangulations quelconques. Pour cela, nous continuerons à explorer dans un premier temps des approches multi-résolutions. Un premier pas vers ces améliorations pourrait être le développement de nouvelles structures de données multi-résolution progressives, notamment des maillages adaptatifs possédant de bonnes propriétés d'invariance topologique locale.

Le premier objectif du projet sera donc le développement et l'évaluation de nouvelles méthodes performantes d'approximation de la topologie des données.

b) Garanties théoriques Un des aspects clefs du calcul progressif est la production de résultats intermédiaires, approchés mais **exploitables**. Ainsi, la qualité de ces résultats intermédiaires a une importance cruciale. Une bonne méthode pour évaluer cette qualité consiste à fournir des garanties théoriques sur les approximations commises tout au long du calcul. Cependant, pour le calcul progressif d'abstractions topologiques, il ne s'agit pas d'une tâche facile en générale. Des métriques bien définies, notamment issues de la théorie du **transport optimal** [37, 46], existent pour les diagrammes de persistance. Elles donnent lieu à des propriétés de stabilité [14, 15, 13] qui peuvent être utilisées pour définir des garanties sur une estimation courante d'un diagramme de persistance, comme je l'ai montré dans mes travaux de thèse [67]. Cependant, pour d'autres signatures topologiques telles que l'arbre de contour, le graphe de Reeb ou le complexe de Morse-Smale, de tels résultats de stabilité n'ont pas encore été démontrés. Pour résoudre ce problème, une direction de recherche pourrait être la définition de nouvelles métriques, dérivées de distances existantes et possédant de bonnes propriétés de stabilité. Une autre piste, dans la lignée de travaux récents, consiste à estimer un niveau de résolution minimum fournissant une bonne approximation, en étudiant la variation locale des données [67] ou à partir d'hypothèse préalables sur la régularité des données [34].

Le second objectif de cet axe de recherche sera ainsi la recherche de garanties fortes sur des approximations de la topologie d'un jeu de données.

Axe 2 : Reconstruction de données à partir d'abstractions topologiques

Motivation

L'analyse topologique de données permet d'encoder des caractéristiques topologiques d'un jeu de données dans des structures de données réduites. Ce processus permet notamment une réduction substantielle de la taille des données, ce qui facilite leur transfert, leur stockage et leur analyse. Cependant, l'opération inverse, c'est-à-dire la reconstruction d'un jeu de donnée à partir d'une signature topologique, a aussi un grand intérêt. On peut voir cette opération comme une étape de **décompression**. Cette reconstruction est encore plus intéressante si on l'applique après modification de la signature topologique considérée, ce qui permet de choisir les caractéristiques topologiques prises en compte dans la reconstruction, et d'effectuer un **débruitage** de la donnée initiale, ou bien d'enlever ou d'ajouter des structures topologiques d'intérêt dans les données. Enfin, on peut envisager une reconstruction à partir d'une signature topologique issue de l'**agrégation** de plusieurs données. J'ai travaillé sur le calcul de telles signatures moyennes, appelées **barycentres**, pendant ma thèse [65, 50]. Ces barycentres peuvent être calculés pour les diagrammes de persistance [63, 65] ou les arbres de jointure [50] à partir d'un jeu de données d'ensemble. La donnée reconstruite à partir de barycentres de signatures topologiques serait donc une donnée "moyenne", représentative de la topologie d'un ensemble de données initial, et pourrait être utilisée à des fins de visualisation ou de comparaison d'ensembles, un sujet d'une importance particulière en visualisation et analyse de données.

Verrous

Il n'existe actuellement pas de méthode canonique pour reconstruire une donnée à partir d'une signature topologique, par exemple un diagramme de persistance. Cela tient au fait que le calcul du diagramme de persistance n'est pas une opération injective (comme illustré en figure 7), notamment car les informations encodées dans le diagramme sont des caractéristiques purement topologiques, indépendantes par exemple de l'échelle ou de l'orientation des données. Ainsi un même diagramme de persistance décrit la topologie d'une infinité de données. Le processus de reconstruction doit donc être aidé par des informations géométriques additionnelles, qui dépendent largement des applications. D'autres abstractions topolgiques comme l'arbre de contour ou le graphe de Reeb, bien qu'encodant plus de structure, souffrent du même problème.

Directions de recherche

a) Définition d'abstractions injectives Une première direction de recherche consiste à définir de nouvelles abstractions topologiques injectives. Des travaux existants montrent qu'il est par exemple possible de caractériser une surface par des *familles* de diagrammes de persistance [64], qui possèdent alors des propriétés d'injectivité.

Le premier objectif de cet axe du projet sera d'étudier plus en détail les propriétés de ces objets et évaluer leur pertinence pour la reconstruction de surfaces, avant d'essayer de généraliser leur usage à des données plus génériques.

b) Reconstruction hiérarchique multi-résolution Mes travaux sur le calcul progressif du diagramme de persistance pour un champ scalaire 2D ou 3D ont montré, en pratique, qu'une approche multi-résolution effectue un bon échantillonnage progressif des structures topologiques des données [66]. La séquence de diagrammes de persistance obtenue à la fin de l'approche encode implicitement des informations sur la répartition géométrique des caractéristiques topologiques des données, en raison de l'échantillonnage qui a été réalisé à chaque niveau de résolution.

Le second objectif de cet axe sera d'étudier comment cette séquence multi-résolution de diagrammes de persistance peut être utilisée pour reconstruire les données initiales, et d'expérimenter plusieurs stratégies

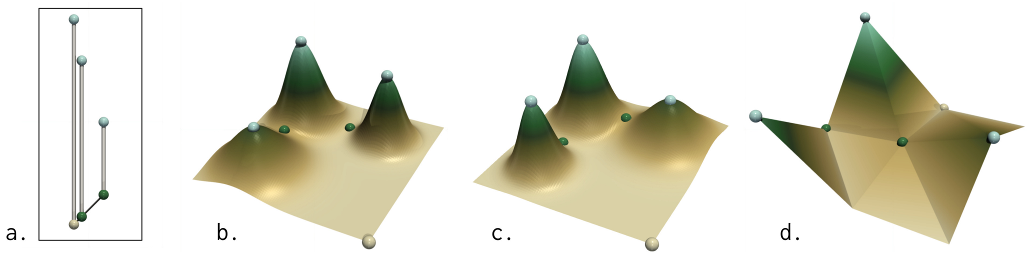

FIGURE 7 – Non-injectivité du diagramme de persistance. Un même diagramme de persistance (a.) décrit correctement la topologie de trois champs scalaires différents (b., c. et d.). En l'absence d'informations géométriques additionnelles, les trois champs constitueraient des reconstructions valides de données à partir du diagramme.

d'échantillonnage. À terme, d'autres abstractions topologiques seront envisagées, comme l'arbre de jointure ou de contour.

c) Ajustement de données Enfin, une approche plus étudiée dans la communauté consiste à ajuster ou compléter des données initiales afin que leur topologie corresponde à une signature topologique cible, notamment par des méthodes d'optimisation différentiable. Des travaux précédents se sont concentrés sur la continuation de nuages de points [25], ou l'optimisation de champs scalaires [51], avec des applications à la correspondance de formes.

Le troisième objectif de cet axe de recherche sera d'étudier ces techniques d'ajustement de données, et particulièrement leur dépendance au choix des données d'entrée.

Axe 3 : Visualisation et intégration dans les applications

Motivation

L'analyse topologique de données est un sujet d'actualité, et bien représenté dans le domaine de la visualisation de donnée [33, 69]. Néanmoins les signatures topologiques calculées en analyse topologique de données sont typiquement utilisées à des fins de comparaison robustes de données et de calculs de correspondances entre jeux de données, ou à des fins d'extraction de structures. Peu de travaux se sont penchés sur l'évaluation de la pertinence de ces signatures en tant qu'outils de visualisation directs, dans le cadre d'une analyse exploratoire des données. En conséquence, ces techniques de visualisation n'ont pas bénéficié d'un transfert vers les domaines d'applications. Pourtant, ces signatures fournissent un résumé visuel explicite de la répartition des structures topologiques présentes dans les données. Les experts en analyse topologique de données sont capables de les lire et d'en déduire rapidement des informations sur la structure des données.

Verrous

Si ces outils de visualisation sont explicites pour des utilisateurs s'y connaissant en analyse topologique de données, ils ne sont pas facilement interprétables en général. J'ai moi-même constaté cette difficulté durant ma thèse, en travaillant sur le projet VESTEC en relation avec des experts des cas d'applications du projet (sous-section 2.3), qui n'avaient pas d'expertise en analyse topologique de données. Cette barrière limite le transfert de ces outils de visualisation vers les domaines d'application.

Directions de recherche

a) Représentation statique Dans un premier temps, l'objectif de cet axe de recherche sera d'explorer de nouvelles représentations visuelles des signatures topologiques, en commençant par des abstractions simples comme le diagramme de persistance ou les arbres de contour. L'impact sur la **perception** de l'utilisateur sera étudié, afin d'identifier des représentations plus à même de transmettre des informations complexes sur la topologie des données. Une première piste de recherche consistera à étudier des techniques du domaine de la visualisation d'information, et des techniques de visualisation illustrative, notamment pour des phénomènes multi-échelles [32].

b) Visualisation dynamique Un second objectif de cet axe de recherche consistera à explorer des représentations dynamiques interactives d'information topologique pour faciliter leur interprétation. Une première direction de recherche consistera à explorer l'impact d'une représentation progressive [70] de diagrammes de persistance sur leur interprétation par des utilisateurs non-experts en analyse topologique de données. Plus tard, d'autres techniques de visualisation interactive issues du domaine de la visualisation d'information pourront être envisagées, comme les techniques de *brushing*.

c) Evaluation Cet axe de recherche devra être soutenu par une évaluation rigoureuse des méthodes de visualisation de signatures topologiques, notamment par le biais d'études utilisateurs. Le déroulement de telles évaluations est un des objectifs principaux de ce troisième axe de mon projet de recherche.

À plus long terme...

Développement d'environnements interactifs pour la visualisation scientifique

Comme expliqué plus haut, j'ai travaillé durant ma thèse sur le développement d'algorithmes progressifs pour l'analyse topologique de données. Ces algorithmes fournissent rapidement des résultats approchés, qui peuvent être interactivement raffinés si plus de précision est demandée par l'utilisateur. Cette progressivité a de multiples intérêts : elle permet d'obtenir un retour visuel interactif durant le calcul, et donc de commencer le processus d'analyse plus tôt, mais aussi potentiellement d'interrompre le calcul ou de modifier des paramètres durant le calcul. Cependant, les environnements logiciels actuellement utilisés en visualisation scientifique se prêtent mal à un tel contrôle interactif de la visualisation.

Dans la suite de mon projet de recherche, à plus long terme, j'aimerais m'impliquer dans le développement d'outils logiciels de visualisation adaptés à la fois à la visualisation et l'analyse topologique de données d'une part, et à l'exploration *interactive* et *progressive* d'autre part. Il s'agit ici de remettre l'interactivité et le contrôle utilisateur au premier plan, devant les capacités d'analyse et de rendu mises en avant dans les logiciels actuels.

Étude et optimisation de l'impact environnemental de méthodes approchées en sciences des données

Outre l'interactivité, les méthodes approchées et progressives en analyse et visualisation de données permettent d'économiser des ressources de calcul. Ainsi, une méthode d'analyse progressive permet d'obtenir le *meilleur* résultat possible dans une contrainte de temps, par exemple. Cet aspect a été mis en avant dans mon travail de thèse, qui s'inscrivait dans le cadre de l'analyse urgente de simulations numériques pour l'aide à la prise de décision en situation de crise. Cependant, la progressivité peut aussi permettre d'économiser d'autres types de ressource.

À mes yeux, ce paradigme est l'occasion d'explorer le développement de méthodes d'analyses de données à faible impact énergétique. L'augmentation continuelle de la puissance des simulateurs et des processus d'acquisition de données entraine une production de données de taille et de complexité toujours croissante, ce qui constitue un défi pour l'analyse interactive de ces données. Face à cela, plutôt que de continuer d'essayer d'augmenter la performance brute des techniques d'analyse de données, on peut faire le choix de se tourner vers des méthodes approchées et progressives. Il s'agit de sacrifier de la précision dans les résultats, au profit d'une certaine économie du temps et de l'énergie dépensés pour le calcul.

Les méthodes topologiques se prêtent particulièrement bien à cette idée, car elles permettent de réduire considérablement les données en les encodant dans des représentations topologiques réduites. De plus, elles fournissent une mesure d'importance multi-échelle sur les structures présentes dans les données. Ces représentations peuvent donc être simplifiées tout en conservant les informations les plus pertinentes pour l'analyse.

À plus long terme, j'aimerais travailler sur l'impact énergétique des méthodes utilisées en sciences des données. L'objectif sera de trouver des moyens de quantifier cet impact, et de comparer l'utilisation de méthodes classiques d'analyse à des méthodes topologiques approchées.

4 Intégration dans l'équipe ORIGAMI du LIRIS

Contacts : Nicolas Bonneel, David Coeurjolly, Guillaume Damiand

Cette section présente mon projet d'intégration au sein de la première unité de recherche demandée, l'équipe ORIGAMI du LIRIS. Mon projet s'insère dans plusieurs thématiques portées au sein de l'équipe : les pôles "Traitement géométrique" et "Perception visuelle et interaction". Des collaborations peuvent également être envisagées avec l'équipe SICAL du LIRIS.

Caractérisation et reconstruction topologique interactive de données médicales

Le projet ANR TOPACS (Traitement Ouvert de données PACS) sur lequel Raphaëlle Chaine et Julie Digne sont partenaires, vise à étudier l'anatomie humaine à travers l'analyse de dizaines de milliers d'images médicales tridimensionnelles fournies par des hôpitaux partenaires. La grande quantité de données à traiter, et le problème de l'**anonymisation** de ces données constituent les principaux défis du projet. Pour résoudre ces problèmes, des méthodes d'extraction de points clés par apprentissage profond sont appliquées. Les objectifs du projet à plus long terme sont la construction d'un modèle complet du corps humain, par agglomération et recalage d'images anonymisées, un problème difficile tant la variabilité inter-individus est importante.

Mes compétences en analyse topologique de données et en analyse d'ensembles de données peuvent me permettre de m'insérer rapidement sur ce projet. En effet, l'encodage des structures topologiques principales dans des signatures réduites permet à la fois de réduire considérablement la taille des données à analyser, et à faire abstraction de la variabilité inter-individus grâce à la notion de persistance topologique.

Les axes 1 et 2 de mon projet de recherche s'articulent bien avec ce projet ANR. Tout d'abord, le calcul interactif de signatures topologiques est nécessaire pour traiter de si grandes quantités de données en des temps raisonnables. Deuxièmement, la reconstruction de donnés à partir de signatures topologiques des images médicales peut apporter des solutions à la fois au problème de l'anonymisation des données, et au problème de fournir un modèle moyen de l'anatomie humaine, après agrégation de signatures topologiques.

Inversement, les domaines de compétence de l'équipe ORIGAMI favorisent des collaborations qui me permettront de mener à bien les objectifs de ces deux axes de recherche.

Les travaux de Guillaume Damiand en analyse topologique de données [2, 17] et surtout sur le calcul approché (avec garanties) de la persistance topologique de maillages par simplification de cartes combinatoires [16] s'inscrivent dans la même direction que mon premier axe de recherche. L'usage de cartes combinatoires, qui m'est pour l'instant étranger, est une piste que je pourrai explorer en collaboration avec l'équipe.

Le second objectif de cet axe consiste à trouver des moyens d'étendre les métriques actuellement utilisées en analyse topologique de données pour dériver des résultats de garanties sur le calcul approché d'informations topologiques. Ces métriques étant pour beaucoup issue de la théorie du transport, l'expertise présente dans l'équipe en transport optimal [56, 49, 40] est un atout nécessaire à la réalisation du projet.

Les travaux de Julie Digne sur la reconstruction de données par transport optimal [21] sont en lien avec la thématique de mon deuxième axe de recherche, bien que la reconstruction s'opère sur des nuages de points et non à partir d'une signature topologique. Une collaboration sur cette thématique permettrait de généraliser ces résultats à une reconstruction plus générique en utilisant des informations topologiques.

Visualisation d'informations topologiques

Des collaborations avec le pôle "Perception visuelle et interaction" de l'équipe peuvent prendre place dans le cadre de mon 3e axe de recherche, la visualisation d'informations topologiques en vue d'une meilleure intégration dans les applications. Johanna Delanoy et Guillaume Lavoué ont notamment travaillé sur l'évaluation de la perception de rendu, à la fois en utilisant des métriques et en effectuant des études utilisateurs [20, 42]. Par ailleurs, une expertise en visualisation interactive et visualisation d'informations est présente dans l'équipe SICAL du LIRIS, avec laquelle des collaborations pourraient être envisagées sur cet axe de recherche, particulièrement avec Romain Vuillemot.

5 Intégration dans l'équipe STORM de l'IRIT

Contacts : Loïc Barthe, Nicolas Mellado, David Vanderhaeghe

Cette section présente mon projet d'intégration dans l'équipe STORM (Structural Models and Tools in Computer Graphics) de l'IRIT, principalement sur la thématique de l'analyse de nuages de points.

Caractérisation globale et multi-échelle de nuages de points par analyse topologique de données

Les avancées des techniques d'acquisition de surface (photogrammétrie, LIDAR) génèrent des données sous forme de nuages de points non-structurés de taille toujours croissante, allant parfois jusqu'à plusieurs millards de points. Une partie des travaux de l'équipe STORM se focalise sur l'analyse de tels nuages de points, afin de reconstruire des informations géométriques, topologiques ou sémantiques sur les surfaces qui ont été échantillonnées. Cette analyse fait face à plusieurs obstacles. Tout d'abord, la complexité des données rend l'analyse difficile. Par nature, les données sont non-structurées, de grande taille, contiennent un certain niveau de bruit dû aux processus d'acquisition, et sont le produit d'un échantillonnage variable. En second lieu, les surfaces à reconstruire possèdent des propriétés d'intérêt à différentes échelles, et l'analyse des nuages de points doit donc être multi-échelle.

Des travaux récents de l'équipe se sont concentrés sur des analyses différentielles de la géométrie de nuages de points *orientés*, pour notamment caractériser la courbure des surfaces échantillonnées [43]. Ces méthodes sont stables et performantes car hautement parallélisables, mais elles caractérisent seulement localement la géométrie des données.

Une autre piste de recherche explorée pendant la thèse de Thibault Lejemble consiste à utiliser l'analyse topologique de données pour caractériser de manière multi-échelle les structures géométriques présentes dans les données. Ces travaux ont montré que la persistance topologique pouvait être utilisée pour caractériser l'importance de structures planaires présentes à différentes échelles dans les données [44]. Ces plans sont reconstruits à de multiples échelles et leurs relations sont encodées dans un graphe de structure, sur lequel l'analyse topologique est effectuée pour caractériser l'évolution des composantes connexes du graphe à différentes échelles. Ces informations topologiques sont utilisées pour analyser interactivement la répartition multi-échelle des composantes, à travers des outils interactifs de sélection, reconstruction et requêtes de correspondances de structures planaires.

Dans la lignée de ce travail, des directions de recherches se dégagent, en relation avec mon projet de recherche. Tout d'abord, on peut envisager d'appliquer directement des techniques d'analyse de données sur le nuage de points. L'avantage d'une telle approche est d'avoir une caractérisation **globale** des structures topologiques des surfaces échantillonnées. Le diagramme de persistance d'un nuage de points peut notamment être calculé à partir de leur complexe de Čech ou de Rips [22]. Cependant cette approche ne passe pas à l'échelle sur des jeux de données contenant des milliards de points, ce qui motive le développement et l'utilisation de techniques approchées et progressives de calcul de signatures topologiques, qui sont au coeur de mon premier axe de recherche. Deuxièmement, l'utilisation de l'analyse topologique de données comme outil interactif d'analyse et de visualisation qui a été faite dans ce travail [44] entre dans le cadre de mon troisième axe de recherche. Il serait intéressant d'étendre cette approche et d'utiliser les signatures topologiques comme outil interactif, en laissant par exemple à l'utilisateur les moyens d'interagir avec le diagramme de persistance.

Couplées aux méthodes de caractérisation géométrique, les méthodes topologiques permettraient d'étendre les travaux précédents sur l'analyse interactive de structures planaires à l'analyse robuste et multi-échelle d'autres structures d'intérêts, notamment des cavités ou des trous bidimensionnels.

Une autre thématique de recherche de l'équipe vise à utiliser l'apprentissage automatique pour des tâches de classification ou de reconnaissance de forme dans les nuages de points. Plutôt que de se tourner vers de larges réseaux de neurones nécessitant beaucoup d'énergie et de temps pour l'entrainement, des récents travaux se sont concentrés sur le développement de nouvelles primitives caractérisant la géométrie ou la topologie des données. Des primitives géométriques locales multi-échelles ont ainsi été utilisées pour l'apprentissage appliqué à la détection d'arêtes dans les nuages de points [35]. Cette description des données permet d'utiliser un réseau de neurones plus léger avec une meilleure performance que les méthodes classiques d'apprentissage profond. Dans la lignée de ces travaux, l'analyse topologique de données pourrait aider à développer de nouvelles primitives topologiques globales pour l'apprentissage automatique.

Étude de l'espace des chemins lumineux d'une scène

Des collaborations ont été aussi envisagées sur l'étude de l'espace des chemins lumineux, c'est-à-dire les trajectoires des rayons lumineux interagissant avec la géométrie et les matériaux d'une scène lors de son rendu.

L'analyse de cet espace est utile au développement de méthodes performantes d'édition de l'illumination d'une scène. Ici, l'analyse topologique de données pourrait être utilisée pour caractériser des chemins d'intérêt, ou bien pour définir des mesures de similarité dans l'espace des chemins. Cette collaboration pourrait prendre place dans le cadre du projet ANR *Structure : hierarchical motion representation for stylized rendering.*

Références

[1] James Ahrens, Berk Geveci, and Charles Law. ParaView : An End-User Tool for Large-Data Visualization. *The Visualization Handbook*, pages 717–731, 2005.

[2] S. Alayrangues, G. Damiand, P. Lienhardt, and S. Peltier. Homology of cellular structures allowing multi-incidence. *Discrete & Computational Geometry (DCG)*, 54(1) :42–77, July 2015.

[3] Keri Anderson, Jeffrey Anderson, Sourabh Palande, and Bei Wang. Topological data analysis of functional MRI connectivity in time and space domains. In *MICCAI Workshop on Connectomics in NeuroImaging*, 2018.

[4] Dimitri P. Bertsekas. A new algorithm for the assignment problem. *Mathematical Programming*, 1981.

[5] H. Bhatia, S. Jadhav, P. Bremer, G. Chen, J. A. Levine, L. G. Nonato, and V. Pascucci. Flow visualization with quantified spatial and temporal errors using edge maps. *IEEE Transactions on Visualization and Computer Graphics*, 2012.

[6] S. Biasotti, D. Giorgio, M. Spagnuolo, and B. Falcidieno. Reeb graphs for shape analysis and applications. *TCS*, 2008.

[7] Talha Bin Masood, Joseph Budin, Martin Falk, Guillaume Favelier, Christoph Garth, Charles Gueunet, Pierre Guillou, Lutz Hofmann, Petar Hristov, Adhitya Kamakshidasan, Christopher Kappe, Pavol Klacansky, Patrick Laurin, Joshua Levine, Jonas Lukasczyk, Daisuke Sakurai, Maxime Soler, Peter Steneteg, Julien Tierny, Will Usher, Jules Vidal, and Michal Wozniak. An Overview of the Topology ToolKit. In *TopoInVis*, 2019.

[8] Alexander Bock, Harish Doraiswamy, Adam Summers, and Cláudio T. Silva. TopoAngler : Interactive Topology-Based Extraction of Fishes. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2018.

[9] R. L. Boyell and H. Ruston. Hybrid techniques for real-time radar simulation. In *Proc. of the IEEE Fall Joint Computer Conference*, 1963.

[10] P.T. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day, and J. Bell. Interactive exploration and analysis of large scale simulations using topology-based data segmentation. *IEEE Transactions on Visualization and Computer Graphics*, 2011.

[11] H. Carr, J. Snoeyink, and U. Axen. Computing contour trees in all dimensions. In *Symp. on Dis. Alg.*, 2000.

[12] Mathieu Carrière, Marco Cuturi, and Steve Oudot. Sliced Wasserstein Kernel for Persistence Diagrams. *ICML*, 2017.

[13] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Oudot. Proximity of persistence modules and their diagrams. In *S. o. C. G.*, 2009.

[14] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *S. o. C. G.*, 2005.

[15] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have lp-stable persistence. *Foundations of Computational Mathematics*, 2010.

[16] G. Damiand, Paluzo-Hidalgo E., R. Slechtac, and R. Gonzalez-Diaz. Approximating lower-star persistence via 2d combinatorial map simplification. *Pattern Recognition Letters (PRL)*, 131 :314–321, March 2020.

[17] G. Damiand and R. Gonzalez-Diaz. Parallel homology computation of meshes. In *Proc. of 6th International Workshop on Computational Topology in Image Context (CTIC)*, volume 9667 of *Lecture Notes in Computer Science*, pages 53–64, Marseille, France, June 2016. Springer International Publishing.

[18] Mark De Berg and Marc van Kreveld. Trekking in the alps without freezing or getting tired. *Algorithmica*, 1997.

[19] Leila De Floriani, Ulderico Fugacci, Federico Iuricich, and Paola Magillo. Morse complexes for shape segmentation and homological analysis : discrete models and algorithms. *Comp. Grap. For.*, 2015.

[20] Johanna Delanoy, Ana Serrano, Belen Masia, and Diego Gutierrez. Perception of material appearance : a comparison between painted and rendered images. *Journal of Vision*, 21(5) :16–16, May 2021.

[21] Julie Digne, David Cohen-Steiner, Pierre Alliez, Fernando de Goes, and Mathieu Desbrun. Feature-preserving surface reconstruction and simplification from defect-laden point sets. *Journal of Mathematical Imaging and Vision*, 48(2) :369–382, 2014.

[22] H. Edelsbrunner and J. Harer. *Computational Topology : An Introduction.* American Mathematical Society, 2009.

[23] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Disc. Compu. Geom.*, 2002.

[24] Jean-Daniel Fekete and Romain Primet. Progressive Analytics : A Computation Paradigm for Exploratory Data Analysis. *arXiv*, 2016.

[25] Marcio Gameiro, Yasuaki Hiraoka, and Ippei Obayashi. Continuation of point clouds via persistence diagrams. *Physica D : Nonlinear Phenomena*, 06 2015.

[26] D. Guenther, R. Alvarez-Boto, J. Contreras-Garcia, J.-P. Piquemal, and J. Tierny. Characterizing molecular interactions in chemical systems. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2014.

[27] A. Gyulassy, P. T. Bremer, B. Hamann, and V. Pascucci. A practical approach to Morse-Smale complex computation : Scalability and generality. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2008.

[28] A. Gyulassy, P.T. Bremer, R. Grout, H. Kolla, J. Chen, and V. Pascucci. Stability of dissipation elements : A case study in combustion. *Comp. Graph. For.*, 2014.

[29] A. Gyulassy, A. Knoll, K.C. Lau, B. Wang, P.T. Bremer, M.E. Papka, L. A. Curtiss, and V. Pascucci. Interstitial and interlayer ion diffusion geometry extraction in graphitic nanosphere battery materials. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2015.

[30] Attila Gyulassy, Peer-Timo Bremer, and Valerio Pascucci. Shared-Memory Parallel Computation of Morse-Smale Complexes with Improved Accuracy. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2018.

[31] Attila Gyulassy, Mark A. Duchaineau, Vijay Natarajan, Valerio Pascucci, Eduardo Bringa, Andrew Higginbotham, and Bernd Hamann. Topologically clean distance fields. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2007.

[32] Sarkis Halladjian, David Kouřil, Haichao Miao, Eduard Gröller, Ivan Viola, and Tobias Isenberg. Multiscale Unfolding : Illustratively Visualizing the Whole Genome at a Glance. *IEEE Transactions on Visualization and Computer Graphics*, March 2021.

[33] C. Heine, H. Leitte, M. Hlawitschka, F. Iuricich, L. De Floriani, G. Scheuermann, H. Hagen, and C. Garth. A survey of topology-based methods in visualization. *Comp. Grap. For.*, 2016.

[34] Teresa Heiss, Sarah Tymochko, Brittany Story, Adélie Garin, Hoa Bui, Bea Bleile, and Vanessa Robins. The impact of changes in resolution on the persistent homology of images, 2021.

[35] Chems-Eddine Himeur, Thibault Lejemble, Thomas Pellegrini, Mathias Paulin, Loic Barthe, and Nicolas Mellado. Pcednet : A lightweight neural network for fast and interactive edge detection in 3d point clouds. *ACM Trans. Graph.*, 41(1), nov 2021.

[36] Jaemin Jo, Jinwook Seo, and Jean-Daniel Fekete. PANENE : A Progressive Algorithm for Indexing and Querying Approximate k-Nearest Neighbors. *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[37] Leonid Kantorovich. On the translocation of masses. *AS URSS*, 1942.

[38] J. Kasten, J. Reininghaus, I. Hotz, and H.C. Hege. Two-dimensional time-dependent vortex regions based on the acceleration magnitude. *IEEE Transactions on Visualization and Computer Graphics*, 2011.

[39] Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. Geometry helps to compare persistence diagrams. *ACM Journal of Experimental Algorithmics*, 2016.

[40] Julien Lacombe, Julie Digne, Nicolas Courty, and Nicolas Bonneel. Learning to generate wasserstein barycenters. *CoRR*, abs/2102.12178, 2021.

[41] D. E. Laney, P.T. Bremer, A. Mascarenhas, P. Miller, and V. Pascucci. Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2006.

[42] Guillaume Lavoué, Nicolas Bonneel, Jean-Philippe Farrugia, and Cyril Soler. Perceptual quality of BRDF approximations : dataset and metrics. *Computer Graphics Forum*, 40(2), May 2021.

[43] Thibault Lejemble, David Coeurjolly, Loïc Barthe, and Nicolas Mellado. Stable and efficient differential estimators on oriented point clouds. *Computer Graphics Forum (Proceedings of Symposium on Geometry Processing)*, 40(5), 2021.

[44] Thibault Lejemble, C Mura, Loic Barthe, and N Mellado. Persistence Analysis of Multi-scale Planar Structure Graph in Point Clouds. *Computer Graphics Forum*, 2020.

[45] Robert B. Miller. Response time in man-computer conversational transactions. In *Fall Joint Computer Conference*, 1968.

[46] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Académie Royale des Sciences de Paris*, 1781.

[47] Malgorzata Olejniczak, André Severo Pereira Gomes, and Julien Tierny. A Topological Data Analysis Perspective on Non-Covalent Interactions in Relativistic Calculations. *International Journal of Quantum Chemistry*, 2019.

[48] V Pascucci, G Scorzelli, P T Bremer, and A Mascarenhas. Robust on-line computation of Reeb graphs : simplicity and speed. *ACM Trans. on Graph.*, 2007.

[49] Loïs Paulin, Nicolas Bonneel, David Coeurjolly, Jean-Claude Iehl, Antoine Webanck, Mathieu Desbrun, and Victor Ostromoukhov. Sliced optimal transport sampling. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, 39(4), July 2020.

[50] Mathieu Pont, Jules Vidal, Julie Delon, and Julien Tierny. Wasserstein Distances, Geodesics and Barycenters of Merge Trees. *IEEE Transactions on Visualization and Computer Graphics*, October 2021.

[51] Adrien Poulenard, Primoz Skraba, and Maks Ovsjanikov. Topological Function Optimization for Continuous Shape Matching. *Computer Graphics Forum*, 37(5) :13–25, 2018.

[52] Georges Reeb. Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *Comptes Rendus des séances de l'Académie des sciences*, 222(847-849) :76, 1946.

[53] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *IEEE CVPR*, 2015.

[54] B. Rieck, F. Sadlo, and H. Leitte. Topological machine learning with persistence indicator functions. In *Proc. of TopoInVis*, 2017.

[55] Vanessa Robins, Peter John Wood, and Adrian P. Sheppard. Theory and Algorithms for Constructing Discrete Morse Complexes from Grayscale Digital Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011.

[56] M. A. Schmitz, M. Heitz, N. Bonneel, F. M. Ngolè Mboula, D. Coeurjolly, M. Cuturi, G. Peyré, and J.-L. Starck. Wasserstein dictionary learning : Optimal transport-based unsupervised non-linear dictionary learning. *SIAM Journal on Imaging Sciences*, 11(1), 2018.

[57] Nithin Shivashankar, Pratyush Pranav, Vijay Natarajan, Rien van de Weygaert, EG Patrick Bos, and Steven Rieder. Felix : A topology based framework for visual exploration of cosmic filaments. *IEEE Transactions on Visualization and Computer Graphics*, 2016. `http://vgl.serc.iisc.ernet.in/felix/index.html`.

[58] Maxime Soler, Martin Petitfrere, Gilles Darche, Melanie Plainchault, Bruno Conche, and Julien Tierny. Ranking Viscous Finger Simulations to an Acquired Ground Truth with Topology-Aware Matchings. In *IEEE Symposium on Large Data Analysis and Visualization*, 2019.

[59] T. Sousbie. The persistent cosmic web and its filamentary structure : Theory and implementations. *Royal Astronomical Society*, 2011. `http://www2.iap.fr/users/sousbie/web/html/indexd41d.html`.

[60] S. Tarasov and M. Vyali. Construction of contour trees in 3d in o(n log n) steps. In *S. o. C. G.*, 1998.

[61] Julien Tierny, Guillaume Favelier, Joshua A. Levine, Charles Gueunet, and Michael Michaux. The Topology ToolKit. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2017. `https://topology-tool-kit.github.io/`.

[62] TTK Contributors. TTK Data. `https://github.com/topology-tool-kit/ttk-data/tree/dev`, 2020.

[63] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet Means for Distributions of Persistence Diagrams. *Disc. Compu. Geom.*, 2014.

[64] Katharine Turner, Sayan Mukherjee, and Doug M. Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference : A Journal of the IMA*, 3(4) :310–344, 2014.

[65] Jules Vidal, Joseph Budin, and Julien Tierny. Progressive wasserstein barycenters of persistence diagrams. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 26(1) :151–161, 2019.

[66] Jules Vidal, Pierre Guillou, and Julien Tierny. A progressive approach to scalar field topology. *IEEE Transactions on Visualization and Computer Graphics*, 27(6) :2833–2850, 2021.

[67] Jules Vidal and Julien Tierny. Fast approximation of persistence diagrams with guarantees. *IEEE Symposium on Large Data Analysis and Visualization*, 2021.

[68] Matt Williams and Tamara Munzner. Steerable, Progressive Multidimensional Scaling. In *Proc. of IEEE InfoVis*, 2004.

[69] Lin Yan, Talha Bin Masood, Raghavendra Sridharamurthy, Farhan Rasheed, Vijay Natarajan, Ingrid Hotz, and Bei Wang. Scalar field comparison with topological descriptors : Properties and applications for scientific visualization. *Comp. Graph. For.*, 2021.

[70] Emanuel Zgraggen, Alex Galakatos, Andrew Crotty, Jean-Daniel Fekete, and Tim Kraska. How progressive visualizations affect exploratory analysis. *IEEE Transactions on Visualization and Computer Graphics*, 2017.