G. de Wit

Persistente homologie en het kosmische web

Bachelorscriptie

Scriptiebegeleiders:
Dr. R.I. van der Veen & Prof.dr. K.H. Kuijken

Datum Bachelorexamen: 4 juli 2016

Mathematisch Instituut & Sterrewacht Leiden, Universiteit Leiden

Inhoudsopgave

1 Introductie

1.1 Opbouw verslag

De structuur van dit verslag is grofweg in drie delen te verdelen. In het eerste deel zullen we de theorie van persistente homologie op filtraties van simpliciale complexen opbouwen. Hierbij beperken we ons tot de voor ons relevante onderwerpen uit eerder werk ([7], [17], [4]) maar zullen we deze op sommige stukken veralgemeniseren en enkele subtiele nalatigheden verbeteren. Verder kijken we naar de computationele kant van persistente homologie, en zullen een algoritme beschrijven dat dit kan uitrekenen.
In het tweede deel passen we de theorie en het algoritme toe op verschillende sterrenkundige datasets. De vraag die ons drijft is of we persisente homologie kunnen gebruiken om te kijken in hoeverre simulaties van het kosmische web overeenkomen met geobserveerde data.
Tot slot evalueren we onze resultaten en kijken naar de mogelijkheden voor vervolgonderzoek.

1.2 Idee achter persistente homologie

Persistente homologie is een tool uit de algebraïsche topologie die gebruikt kan worden om data te analyseren. Wij zullen ons beperken tot het analyseren van zogenaamde “point cloud data”, maar de theorie vindt ook andere toepassingen.
Van een eindige verzameling $S \subseteq \mathbb{R}^n$, een *puntenwolk*, maken we een simpliciaal complex door punten die “dicht genoeg” bij elkaar liggen te verenigen tot een simplex. Op die manier proberen we de achterliggende topologische ruimte waar de puntenwolk van afkomstig is te benaderen. De vraag is daarbij wat hierbij “dicht genoeg” is. Dit lossen we op door naar een breed bereik te kijken van afstanden, en voor elk van die afstanden een dergelijk simpliciaal complex te bouwen. Op deze manier kunnen we kijken naar op welk gebied van afstanden bepaalde topologische eigenschappen bestaan (denk bijvoorbeeld aan gaten en samenhangscomponenten). Topologische eigenschappen met een kort bereik kunnen gekarakteriseerd worden als ruis, terwijl eigenschappen met een groter bereik met meer zekerheid topologische eigenschappen van de onderliggende ruimte zijn.

Deel I
Theorie

Conventies

We houden onder andere de volgende conventies en notaties aan:

- $\mathbb{N} = \mathbb{Z}_{\geq 0}$;
- $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$;
- $\infty - n = \infty$ voor alle $n \in \mathbb{N}$.

2 Voorkennis

De lezer wordt bekend verondersteld met de volgende algebraïsche structuren: groepen, ringen, lichamen en modulen. Een introductie in deze structuren kan gevonden worden in de dictaten van P. Stevenhagen ([15], [14]).

2.1 Simpliciale complexen

Aan de basis van persistente homologie ligt simpliciale homologie. Om dit te kunnen introduceren, specificeren we eerst de door ons gebruikte versie van een simplicaal complex.

Definitie 2.1 (Simpliciaal complex)**.** Een *simpliciaal complex* is een *puntenverzameling* S, samen met een een eindige, niet-lege deelverzameling $\Delta \subseteq \mathcal{P}(S) \setminus \{\varnothing\}$ van eindige, niet-lege deelverzamelingen van S, met de volgende eigenschap:

$$\forall \omega \in \Delta : \{\tau \subseteq \omega : \tau \neq \varnothing\} \subseteq \Delta.$$

We noteren een simplicaal complex als (Δ, S), en zeggen dat Δ *een simplicaal complex is op* S, en we schrijven kortweg Δ, als uit de context duidelijk is wat S is.

Zij (Δ, S) een simplicaal complex. Een element $\omega \in \Delta$ heet een *k-simplex* als $|\omega| = k + 1$; we zeggen ook dat ω *dimensie* k heeft. Een element $\omega \supsetneq \tau \in \Delta$ heet een *zijde* van ω.

Definitie 2.2 (oriëntatie)**.** Zij Δ een simpliciaal complex op S en $\omega = \{v_0, v_1, ..., v_k\} \in \Delta$ een k-simplex. Een *oriëntatie* op ω is een equivalentieklasse van ordeningen van de punten van ω, waarbij voor $\varphi \in S_k$ geldt: $(v_0, v_1, ..., v_k) \sim (v_{\varphi(0)}, ..., v_{\varphi(1)}, v_{\varphi(k)})$ dan, en slechts dan als $\operatorname{sgn}(\varphi) = 1$. Notatie: we schrijven $[\omega]$ voor een simplex met oriëntatie; een *georiënteerde simplex*. Voor de geörienteerde k-simplex $[\omega]$, schrijven we ook $[\omega] = [v_0, ..., v_k]$, waarbij dus geldt

$$[v_0, ..., v_k] = [v_{\varphi(0)}, ..., v_{\varphi(n)}] \Leftrightarrow \varphi \text{ is een even permutatie.}$$

In onze toepassing zullen we gebruik maken van simpliciale complexen die voortkomen uit een eindige deelverzameling punten van $\mathbb{R}^n$: het zogenaamde *Vietoris-Rips complex*.

Definitie 2.3 (Vietoris-Rips complex)**.** Zij $\varepsilon \in \mathbb{R}_{\geq 0}$. Het *Vietoris-Rips ε-complex*, kortweg *VR-complex* van graad $n \geq 1$, behorende bij een eindige niet-lege puntenverzameling $S \subseteq \mathbb{R}^n$, is het simpliciale complex $\Delta_\varepsilon(S)$, verkregen door alle deelverzamelingen van S met maximaal $n + 1$ elementen te nemen die paarsgewijs op afstand hoogstens 2ε van elkaar liggen. Hierbij noemen we ε de *koppellengte* van het VR-complex.

2.2 Simpliciale homologie

We nemen vanaf nu aan dat R een hoofdideaaldomein is met eenheidselement. Om naar simpliciale homologie toe te werken, definiëren we eerst een algebraïsche structuur op een complex.

Definitie 2.4 (k-ketens)**.** Zij Δ een simpliciaal complex en $k \in \mathbb{N}$. De verzameling $C_k(\Delta)$ van *k-ketens op* Δ is het vrije R-moduul, voortgebracht door de verzameling van georiënteerde k-simplices van Δ, met $[\omega] = -[\tau]$ als σ en τ hetzelfde k-simplex zijn, met verschillende oriëntatie. Als het uit de context duidelijk is wat Δ is, schrijven we kortweg C_k voor de k-ketens van Δ.

Voor een georiënteerde simplex kunnen we ook aangeven wat de rand ervan is.

Definitie 2.5 (rand)**.** Zij Δ een simpliciaal complex, $k \in \mathbb{N}, k > 0$. Voor een georiënteerde k-simplex $[\omega] = [v_0, ..., v_k] \in \Delta$ is de *rand* van $[\omega]$ gegeven door:

$$\partial_k([\omega]) = \sum_{i=0}^{k} (-1)^i [v_0, ..., \hat{v}_i, ..., v_k],$$

waarbij $[v_0, ..., \hat{v}_i, ..., v_k]$ de georiënteerde $(k-1)$-zijde van ω is, waarbij v_i is weggelaten.

Zetten we ∂_k, voor $k > 0$, R-lineair voort op $C_k(\Delta)$, dan krijgen we een R-homomorfisme $\partial_k \colon C_k \to C_{k-1}$; de *$k$-de randafbeelding*. Schrijven we verder $\partial_0 \colon C_0 \to 0$ als het triviale R-homomorfisme, dan krijgen we de volgende rij van R-homomorfismen:

$$\cdots \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \longrightarrow \cdots \longrightarrow C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0.$$

Lemma 2.6. *Zij Δ een simpliciaal complex. Voor alle $k \in \mathbb{N}$ geldt $\partial_k \partial_{k+1} = 0$.*

Bewijs. Dit volgt direct uit lemma 2.1 van Hatcher [8]. □

Lemma 2.6 is de fundamentele eigenschap die het zinnig maakt om naar homologie te kijken.

Definitie 2.7 (k-cycli en -grenzen)**.** Laat Δ een simpliciaal complex zijn. Voor iedere $k \in \mathbb{N}$ definiëren we de volgende deelmodulen van C_k:

$$Z_k(\Delta) := \ker \partial_k$$
$$B_k(\Delta) := \mathrm{im}\partial_{k+1}$$

We noemen $Z_k(\Delta)$ de verzameling van *k-cycli* van Δ, en $B_k(\Delta)$ de verzameling van *k-grenzen*. Ook hier schrijven we Z_k en B_k, als het duidelijk is om welke Δ het gaat.

Merk op dat lemma 2.6 garandeert dat $\mathrm{im}\partial_{k+1} \subseteq \ker \partial_k$. We krijgen dus de inclusies $B_k \subseteq Z_k \subseteq C_k$ van R-modulen. We hebben nu al het materiaal in handen om homologie te kunnen definiëren:

Definitie 2.8 (Homologie)**.** Voor een simpliciaal complex Δ definiëren we de *k-de homologie* van Δ, $H_k(\Delta)$, als volgt:

$$H_k(\Delta) := Z_k / B_k.$$

Het *k-de Betti-getal* definiëren we als de vrije rang van $H_k(\Delta)$, notatie: $\beta_k(\Delta) := \mathrm{rang} H_k(\Delta)$. Ook bij de k-de homologiën en Betti-getallen laten we de Δ weg, als er geen verwarring kan ontstaan.

De elementen $\sigma \in H_k$ heten *k-de, of k-dimensionale homologieklassen*, of kortweg *klassen* en zullen we noteren met Griekse letters. In sommige gevallen willen we expliciet met een representant uit Z_k werken, en zullen $\overline{\sigma} = \sigma + B_k \in H_k$ schrijven.
Tot slot kijken we kort nog naar afbeeldingen tussen simpliciale complexen.

Definitie 2.9 (simpliciale afbeelding)**.** Een *simpliciale afbeelding* van (Δ_S, S) naar (Δ_T, T) is een afbeelding $f\colon S \to T$, zodanig dat voor alle $\sigma \in \Delta_S$ geldt $f(\sigma) \in \Delta_T$. We noteren een simpliciale afbeelding vaak als $f\colon \Delta_S \to \Delta_T$, waarbij f impliciet gedefinieerd is van S naar T.

Een simpliciale afbeelding $f\colon \Delta_S \to \Delta_T$ geeft op de volgende manier een afbeelding $f_\sharp\colon C_k(\Delta_S) \to C_k(\Delta_T)$,

$$[v_0, ..., v_k] \mapsto \begin{cases} [f(v_0), ..., f(v_k)], & \text{als } f(v_0), ..., f(v_k) \text{ allemaal onderling verschillend} \\ 0, & \text{anders,} \end{cases}$$

en dit zetten we R-lineair voort op $C_k(\Delta_S)$. In Hatcher [8] is te lezen dat een dergelijke afbeelding, voortgekomen uit een simpliciale afbeelding een *ketenmorfisme* is. Zonder hier verder op de details in te gaan, merken we op dat $f_\sharp$ tot een R-homomorfisme $f_*\colon H_k(\Delta_S) \to H_k(\Delta_T)$ leidt, wat betekent dat het nemen van homologie een functor is van de categorie van simpliciale complexen met simpliciale afbeeldingen, naar de categorie van R-modulen met R-homomorfismen.

3 Persistente homologie

Om persistente homologie te kunnen toepassen, moeten we eerst het raamwerk opbouwen waarin de theorie bedreven wordt. In de rest van deze sectie veronderstellen we dat homologie over een hoofdideaaldomein R wordt genomen, tenzij anders vermeld.

3.1 Persistente homologie

Definitie 3.1 (filtratie)**.** Een *filtratie van simpliciale complexen* op een gegeven puntenverzameling S is een keten

$$\Delta^0 \subseteq \Delta^1 \subseteq \cdots$$

van genestelde simpliciale complexen op S, die we noteren als $\mathcal{F} := (\Delta^i)_{i\in\mathbb{N}}$. Een filtratie heet *eindig* als de keten stabiliseert; $\Delta^i = \Delta^n$, voor alle i groter of gelijk aan een zekere $n \in \mathbb{N}$. In dit laatste geval zeggen we ook dat $\mathcal{F}$ een *filtratie is van* Δ^n.

Gegeven een filtratie $\mathcal{F}$, schrijven we $\Omega(\mathcal{F}) := \bigcup_{i\in\mathbb{N}} \Delta^i$, of kortweg Ω, als uit de context duidelijk is om welke filtratie het gaat. Verder noteren we voor $\omega \in \Omega$ het volgende: $\mathrm{in}(\omega) = \min_{i\in\mathbb{N}}\{\Delta^i : \omega \in \Delta^i\}$.

Voorbeeld 3.2 (VR-filtratie)**.** Zij $S \subseteq \mathbb{R}^n$ een eindige niet-lege deelverzameling, voor zekere $n \geq 1$. Laat $(\varepsilon_i)_{i\in\mathbb{N}}$ een niet-dalende reële rij zijn. Gebruik makende van definitie 2.3, vinden we de volgende *VR-filtratie*: $F_{\mathrm{VR}}(S) := (\Delta_{\varepsilon_i}(S))_{i\in\mathbb{N}}$. Dit is inderdaad een filtratie, aangezien voor $j > i$ geldt $\varepsilon_j \geq \varepsilon_i$, dus volgt direct uit definitie 2.3 $\Delta_{\varepsilon_i}(S) \subseteq \Delta_{\varepsilon_j}(S)$. Merk op dat S eindig is, dus $\mathcal{F}_{\mathrm{VR}}$ is ook eindig. In onze toepassing zullen we de persistente homologie van VR-filtraties berekenen.

Sommige eigenschappen laten zich makkelijker beschrijven aan de hand van een *fijne* filtratie.

Definitie 3.3 (fijne filtratie)**.** Een filtratie van simpliciale complexen $\mathcal{F}$ heet *fijn*, als voor alle $i \in \mathbb{N}$ geldt $\#(\Delta^{i+1} \setminus \Delta^i) \in \{0, 1\}$.

Om eigenschappen van een filtratie te kunnen definiëren aan de hand van fijne filtraties, beschouwen we ook het volgende begrip.

Definitie 3.4 (verfijning)**.** Gegeven een filtratie $\mathcal{F}$ op S, zeggen we dat een filtratie $\tilde{\mathcal{F}} = (\tilde{\Delta}^i)_{i\in\mathbb{N}}$ een *verfijning* is van $\mathcal{F}$, als aan elk van de volgende eigenschappen wordt voldaan:

(i) $\tilde{\mathcal{F}}$ is een fijne filtratie op S;

(ii) Er geldt $\Omega(\tilde{\mathcal{F}}) = \Omega(\mathcal{F})$;

(iii) Er is een niet-dalende rij $(n_j)_{j\in\mathbb{N}}$ zodanig dat geldt $\tilde{\Delta}^{n_j} = \Delta^j$.

Het idee achter een verfijning is dat de simplices van elke Δ^i in een filtratie, stuk voor stuk worden toegevoegd.
Bij een filtratie $\mathcal{F}$ krijgen we natuurlijke inclusies $\iota^{i,j} : \Delta^i \to \Delta^j$, voor $i \leq j$. We noteren $H_k^i := H_k(\Delta^i)$ voor de k-de homologie op Δ^i, met $i, k \in \mathbb{N}$, en analoog schrijven we C_k^i, Z_k^i en B_k^i voor de (k-dimensionale) ketens, cycli en grenzen van Δ^i respectievelijk. De filtratie levert op deze manier een keten van homomorfismen $H_k^0 \to H_k^1 \to \cdots$, geïnduceerd door de inclusies. Voor $i, j \in \mathbb{N}$ met $i \leq j$ schrijven we $f_k^{i,j} : H_k^i \to H_k^j$ voor het homomorfisme geïnduceerd door $\iota^{i,j}$.
Voor een verfijning $\tilde{\mathcal{F}}$ van $\mathcal{F}$ schrijven we $\tilde{H}_k^i := H_k(\tilde{\mathcal{F}})$. Het volgende lemma geeft het verband tussen een filtratie en een verfijning ervan goed weer:

Lemma 3.5. *Zij $\mathcal{F}$ een filtratie met verfijning $\tilde{\mathcal{F}}$, $j, n_j \in \mathbb{N}$ zodanig dat $\tilde{\Delta}^{n_j} = \tilde{\Delta}^j$. Voor iedere $k \in \mathbb{N}$ is er een canoniek R-isomorfisme $\eta_k^{j,n_j} : H_k^j \to \tilde{H}_k^{n_j}$.*

Bewijs. Merk op dat de complexen $\tilde{\Delta}^{n_j}$ en Δ^j precies dezelfde simplices bevatten. De canonieke simpliciale afbeelding $i\colon \Delta^j \to \tilde{\Delta}^{n_j}, \sigma \to \sigma$ levert een R-isomorfisme $\eta_k^{j,n_j} : H_k^j \to \tilde{H}_k^{n_j}$. □

Voor persistente topologische eigenschappen zijn we geïnteresseerd in de volgende definitie.

Definitie 3.6 (persistente homologie)**.** Zij $i, j, k \in \mathbb{N}$ met $i \leq j$ en $\mathcal{F}$ een filtratie van simpliciale complexen. Dan is de *j-persistente k-de homologie van Δ^i* gegeven door $H_k^{i,j} := \mathrm{im} f_k^{i,j}$. Analoog definiëren we het *j-persistente k-de Betti-getal* als $\beta_k^{i,j} := \mathrm{rang} H_k^{i,j}$.

Analoog aan wat we eerder deden, schrijven we voor een verfijning $\tilde{\mathcal{F}}$ van $\mathcal{F}$, $\tilde{f}_k^{i,j}\colon \tilde{H}_k^i \to \tilde{H}_k^j$ voor de afbeelding verkregen door de inclusie $\tilde{\iota}^{i,j}\colon \tilde{\Delta}^i \to \tilde{\Delta}^j$, evenals $\tilde{H}_k^{i,j} := \mathrm{im}\tilde{f}_k^{i,j}$.
De persistente homologie kunnen we ook op een andere manier karakteriseren:

Lemma 3.7. *Zij $i, j, k \in \mathbb{N}$ met $i \leq j$ en $\mathcal{F}$ een filtratie van simpliciale complexen. Dan geldt $H_k^{i,j} \cong Z_k^i/(B_k^j \cap Z_k^i)$.*

Bewijs. Allereerst merken we op dat B_k^j en Z_k^i beiden deelmodulen van C_k^j zijn (onder identificatie van Z_k^i met $\iota_\sharp^{i,j}(Z_k^i)$), dus $B_k^j \cap Z_k^i$ is een deelmoduul van Z_k^i, dus $Z_k^i/(B_k^j \cap Z_k^i)$ is inderdaad een R-moduul. Bekijk nu de volgende afbeelding:

$$\begin{aligned}\varphi\colon Z_k^i/(B_k^j \cap Z_k^i) &\to H_k^{i,j},\\ \sigma + B_k^j \cap Z_k^i &\mapsto f_k^{i,j}(\sigma + B_k^i) = \iota_\sharp^{i,j}(\sigma) + B_k^j.\end{aligned}$$

Zij $\sigma, \sigma' \in Z_k^i$ zodanig dat $\sigma - \sigma' \in B_k^j \cap Z_k^i$. Merk op dat geldt $\iota_\sharp^{i,j}(\sigma) - \iota_\sharp^{i,j}(\sigma') = \iota_\sharp^{i,j}(\sigma - \sigma') \in B_k^j$, dus $\iota_\sharp^{i,j}(\sigma) = \iota_\sharp^{i,j}(\sigma')$, waaruit direct volgt dat φ welgedefinieerd is. Daarnaast volgt dat φ een R-homomorfisme is direct uit het feit dat $\iota_\sharp^{i,j}$ dat is. Er rest slechts te bewijzen dat φ een bijectie is. Stel dat $\varphi(\overline{\sigma}) = \overline{0}$, voor een zekere $\sigma \in Z_k^i$. Dan geldt dus $\iota_\sharp^{i,j}(\sigma) \in B_k^j$, en aangezien $\iota_\sharp^{i,j}$ een inclusie is, volgt direct dat geldt $\sigma \in Z_k^i \cap B_k^j$, dus we concluderen dat $\ker\varphi = \overline{0}$. Laten we tot slot $\overline{\sigma'} \in H_k^{i,j}$ willekeurig, dan bestaat er een $\sigma \in Z_k^i$ met $f_k^{i,j}(\sigma + B_k^i) = \iota_\sharp^{i,j}(\sigma) + B_k^j = \overline{\sigma'}$, dus dan geldt ook $\varphi(\sigma + B_k^j \cap Z_k^i) = \overline{\sigma}$, waaruit volgt dat φ surjectief is. Conclusie: φ is een R-moduulisomorfisme. □

Informeel vertelt lemma 3.7 ons dat $H_k^{i,j}$ de klassen van H_k^i zijn, die nog niet triviaal zijn geworden in H_k^j. We gaan nu toewerken naar een maat die aangeeft hoe 'persistent' klassen in een filtratie zijn.

3.2 Persistentie-intervallen en -diagrammen

Een klasse $\sigma \in H_k^i$ heet *geboren in* Δ^i, als $i = 0$, of als $\sigma \notin H_k^{i-1,i}$. Hebben we $0 \leq i < j$ en $\sigma \in H_k^j$, met $(f_k^{i,j})^{-1}(\sigma) \neq \varnothing$, dan noemen we een klasse $\sigma' \in (f_k^{i,j})^{-1}(\sigma)$ een *voorouder* van σ. Als we de kleinste $i \leq j$ nemen waarvoor $(f_k^{i,j})^{-1}(\sigma)$ niet leeg is, dan noemen we de elementen van $(f_k^{i,j})^{-1}(\sigma)$ *stamouders* van σ, en in het algemeen noemen we een klasse $\sigma' \in H_k^{i'}$ een stamouder als er een $i \geq i'$ en $\sigma \in H_k^i$ zijn waarvoor σ' een stamouder is. Merk hierbij op dat elke klasse minstens één stamouder heeft, en dat elke stamouder een stamouder van zichzelf is.

Propositie 3.8. *Zij $\mathcal{F}$ een filtratie, $i, k \in \mathbb{N}$ en $\sigma \in H_k^i$. De volgende uitspraken zijn equivalent:*

(i) σ is geboren in Δ^i.

(ii) σ heeft geen voorouders.

(iii) σ is een stamouder (van zichzelf).

Bewijs. We bewijzen de equivalentie van de uitspraken door 3 implicaties te bewijzen:

$(i) \Rightarrow (ii)$ Neem aan dat σ geboren is in Δ^i. Als $i = 0$, volgt direct dat σ geen voorouders heeft. Neem nu aan dat $i > 0$ en merk op dat geldt $(f_k^{i-1,i})^{-1}(\sigma) = \varnothing$, want $\sigma \notin H_k^{i-1,i}$. Veronderstel dat er een $l < i$ is met $(f_k^{l,i})^{-1}(\sigma) \neq \varnothing$. Dan volgt $f_k^{i-1,i}(f_k^{l,i}(H_k^l)) \ni \sigma$, wat een tegenspraak oplevert, dus een dergelijke l bestaat niet. We concluderen dat σ geen voorouders heeft.

$(ii) \Rightarrow (iii)$ Neem aan dat σ geen voorouders heeft. Als $i = 0$, dan volgt direct uit de definitie dat σ een stamouder is. Als $i > 0$ weten we dat $(f_k^{l,i})^{-1}(\sigma) = \varnothing$ voor alle $l < i$, en we hebben natuurlijk $\sigma \in (f_k^{i,i})^{-1}(\sigma)$, dus we concluderen dat σ een stamouder is.

$(iii) \Rightarrow (i)$ Neem aan dat σ een stamouder is. Voor $i = 0$ geldt automatisch dat σ geboren is in Δ^i. Voor $i > 0$ weten we dat $(f_k^{l,i})^{-1}(\sigma)$ leeg is voor alle $l < i$, dus in het bijzonder geldt $\sigma \notin H_k^{i-1,i}$. Conclusie: σ is geboren in Δ^i.

Uit de drie implicaties volgt de equivalentie van de uitspraken $(i), (ii)$ en (iii). □

We zeggen dat een stamouder $\sigma \in H_k^i$ *ouder* is dan een stamouder $\sigma' \in H_k^{i'}$, als $i < i'$.
Voor een klasse $\sigma \in H_k^i$ die geboren is in Δ^i, willen we ook kunnen zeggen dat deze sterft, indien deze in een zekere Δ^j samengaat met een oudere stamouder. Dit ligt alleen iets gecompliceerder in het geval twee klassen samengaan die in dezelfde Δ^i zijn geboren. Om dit op te lossen geven we een constructie voor een verfijning van $\mathcal{F}$.

Definitie 3.9 (compatibele ordening)**.** Een *compatibele ordening* op een filtratie $\mathcal{F}$ is een injectieve afbeelding $\mathrm{ord}\colon \Omega(\mathcal{F}) \to \mathbb{N}$, waarvoor zowel de implicatie $\mathrm{in}(\omega) < \mathrm{in}(\omega') \Rightarrow \mathrm{ord}(\omega) < \mathrm{ord}(\omega')$ als $\omega \subseteq \omega' \Rightarrow \mathrm{ord}(\omega) \leq \mathrm{ord}(\omega')$ geldt, en waarvoor geldt $0 \in \mathrm{im}(\mathrm{ord})$.

Voor een filtratie $\mathcal{F}$ met compatibele ordening ord schrijven we voor elke $i \in \mathbb{N}$, $\mathrm{ord}(\Delta^i) := \max\{\mathrm{ord}(\omega) : \omega \in \Delta^i\}$.

Lemma 3.10. *Voor elke filtratie $\mathcal{F}$ bestaat een compatibele ordening.*

Bewijsschets. We geven een intuïtief duidelijke constructie van hoe een compatibele ordening kan worden geconstrueerd voor een gegeven filtratie $\mathcal{F}$. Merk op dat er maar eindig veel elementen $\omega \in \Omega$ zijn met $\mathrm{in}(\omega) = i$, voor alle $i \in \mathbb{N}$. We kijken naar de verzamelingen $\mathrm{in}_i(\Omega) := \{\omega \in \Omega : \mathrm{in}(\omega) = i\}$. Schrijf $n_i = \#\mathrm{in}_i(\Omega)$. Het is duidelijk dat we een injectieve afbeelding $\mathrm{ord}_i\colon \mathrm{in}_i(\Omega) \to \{1, ..., n_i\}$ kunnen maken, zodanig dat voor $\tau \subseteq \omega \in \mathrm{in}_i(\Omega)$ geldt $\mathrm{ord}_i(\tau) < \mathrm{ord}_i(\omega)$, waarbij ord_i

de triviale afbeelding $\varnothing \to \{0\}$ is, als $n_i = 0$. Een compatibele ordening ord kan nu als volgt worden geconstrueerd:

$$\mathrm{ord}(\omega) = \left(\mathrm{ord}_{\mathrm{in}(\omega)}(\omega) + \sum_{i=0}^{\mathrm{in}(\omega)} n_i \right) - 1,$$

waarbij de -1 garandeert dat geldt $0 \in \mathrm{im}(\mathrm{ord})$ (aangezien er minstens 1 element is in $\mathrm{ord}_0(\Omega)$, want complexen zijn niet leeg). □

De eis dat 0 in het beeld van ord zit, is van technische aard en zorgt ervoor dat de simpliciale complexen in de volgende definitie niet leeg zijn.

Definitie 3.11 (geordende filtratie)**.** Een *geordende filtratie* behorende bij een filtratie $\mathcal{F}$ met compatibele ordening ord, is de keten $\tilde{\mathcal{F}} := (\tilde{\Delta}^n)_{n\in\mathbb{N}}$, met

$$\tilde{\Delta}^n = \{\omega \in \Omega((\Delta)^i_{i\in\mathbb{N}}) : \mathrm{ord}(\omega) \leq n\}.$$

Opmerking 3.12. Een geordende filtratie behorende bij $(\mathcal{F}, \mathrm{ord})$ is een verfijning van $\mathcal{F}$. Dit volgt direct uit de definitie van ord, waarbij we opmerken dat we als niet-dalende rij $n_j = \mathrm{ord}(\Delta^j)$ kunnen nemen, voor alle $j \in \mathbb{N}$. In het bijzonder volgt uit lemma 3.10 dat er voor elke filtratie een verfijning bestaat. Daarnaast levert een verfijning $\tilde{\mathcal{F}}$ van $\mathcal{F}$ op een natuurlijke manier een ordening; $\mathrm{ord}\colon \Omega(\tilde{\mathcal{F}}) \to \mathbb{N}, \omega \mapsto \mathrm{in}(\omega)$. Om die reden gebruiken we beide begrippen door elkaar met dezelfde notatie $\tilde{\mathcal{F}}$.

Merk op dat compatibele ordeningen in het algemeen niet uniek zijn, en er dus meerdere mogelijke geordende filtraties zijn. Later zullen we zien dat de keuze van een compatibele ordening niet uitmaakt voor de voor ons interessante eigenschappen van persistente homologie.
De definitie van uitsterven van een stamouder $\sigma \in H^i_k$ is vrij gemakkelijk in termen van fijne filtraties.

Definitie 3.13 (uitsterven in fijne filtratie)**.** Laat $\mathcal{F}$ een fijne filtratie zijn en $\sigma \in H^i_k$ een stamouder van zichzelf, met $i > 0$. We zeggen dat de klasse σ *uitsterft* in Δ^j, met $j > i$, als:

$$f^{i,j-1}_k(\sigma) \notin H^{i-1,j-1}_k, \text{ en}$$
$$f^{i,j}_k(\sigma) \in H^{i-1,j}_k.$$

Deze definitie maakt het samengaan van een stamouder met een andere stamouder formeel. Als een stamouder $\sigma \in H^i_k$ van een fijne filtratie uitsterft in Δ^j, zorgt de constructie ervoor dat σ samengaat met een oudere stamouder σ'. Hierbij sterft σ' dus niet uit.
In een algemene filtratie $\mathcal{F}$ kunnen verschillende stamouders tegelijk geboren worden en in dezelfde Δ^j uitsterven. Hier moet dus een keuze gemaakt worden welke de rol van oudere op zich neemt, en welke klasse uitsterft. Deze keuze wordt voor ons gemaakt aan de hand van een compatibele ordening.

Definitie 3.14 (uitsterven)**.** Zij $\mathcal{F}$ een filtratie met compatibele ordening ord, $i, k \in \mathbb{N}$ en $\sigma \in H^i_k$ een klasse die geboren is in Δ^i. Schrijf $l := \mathrm{ord}(\Delta^n)$ en $\sigma' := \eta^{n,l}_k(\sigma)$ en laat σ'_s een stamouder zijn van σ', geboren in $\tilde{\Delta}^{i'}$ en neem aan dat geldt $i' > 0$. We zeggen dat σ *uitsterft in* Δ^j, met $j > i$ als er een $\omega \in \Delta^j$ is waarvoor geldt dat σ'_s uitsterft in $\tilde{\Delta}^{\mathrm{ord}(\omega)}$. Indien er geen enkele $j > i$ bestaat waarvoor geldt dat σ uitsterft in Δ^j, zeggen we dat σ *nooit uitsterft*, ofwel *oneindig blijft leven*. In het geval dat σ'_s geboren wordt in $\tilde{\Delta}^0$, zeggen we ook dat σ nooit uitsterft.

Voor een stamouder $\sigma \in H_k^i$ schrijven we:

$$\text{sterfte}(\sigma) = \begin{cases} j, & \text{als } \sigma \text{ uitsterft in } \Delta^j \\ \infty, & \text{als } \sigma \text{ nooit uitsterft.} \end{cases}$$

Intuïtief willen we nu zeggen dat hoe langer klassen blijven leven, hoe 'persistenter' ze zijn. De volgende definitie geeft een maat voor hoe lang klassen leven.

Definitie 3.15 (persistentie(-interval))**.** Zij $(\mathcal{F}, \text{ord})$ een filtratie met compatibele ordening, $i, k \in \mathbb{N}$ en $\sigma \in H_k^i$ een stamouder. Het *persistentie-interval* van σ is het paar $\text{int}(\sigma) := (i, \text{sterfte}(\sigma)) \in \mathbb{N} \times \overline{\mathbb{N}}$. De *persistentie* van σ definiëren we als $\text{pers}(\sigma) := \text{sterfte}(\sigma) - i$, ook wel de *lengte* van het persistentie-interval van σ.

Hebben we een filtratie $\mathcal{F}$ met verfijning $\tilde{\mathcal{F}}$, dan kunnen we voor een stamouder $\sigma' \in \tilde{H}_k^{i'}$, met persistentie-interval $\text{int}(\sigma') = (i', j')$, vrij gemakkelijk zien of dit correspondeert met een persistentie-interval van $\mathcal{F}$. Er zijn namelijk unieke simplices $\omega \in \Delta^{i'}$ en $\omega' \in \Delta^{j'}$ met $\tilde{\text{in}}(\omega) = i'$ en $\tilde{\text{in}}(\omega') = j'$. Bekijken we deze simplices als elementen van Ω, dan schrijven we $i = \text{in}(\omega)$ en $j = \text{in}(\omega')$. Als geldt $j > i$, dan volgt dat (i, j) een persistentie-interval is, behorende bij een zekere klasse $\sigma \in H_k^i$, namelijk die waarvoor geldt $\eta_k^{i,\text{ord}(\Delta^i)}(\sigma) = \sigma'$. Deze procedure gaan we later gebruiken om via de computer persistente homologie te berekenen.
We willen een manier hebben om persistente homologie grafisch weer te geven. Hiervoor maken we gebruik van het begrip *persistentie-diagram*. Voordat we daar de definitie van geven, voeren we nog de volgende notatie in voor $i, j \in \mathbb{N} \cup \{-1\}$ met $j > i$, en dimensie $k \in \mathbb{N}$:

$$N_k^{i,j} = \begin{cases} 0, & \text{als } i = -1 \\ \beta_k^{i,j-1} - \beta_k^{i,j}, & \text{anders.} \end{cases}$$

Definitie 3.16 (Persistentie-diagram)**.** Het *persistentie-diagram* $\text{Dgm}_k(\mathcal{F})$ van de filtratie met compatibele ordening $(\mathcal{F}, \text{ord})$ is een multiset, met elementen van de vorm $(i, j) \in \mathbb{N} \times \overline{\mathbb{N}}$ met $j > i$, waarbij een dergelijk paar (i, j) multipliciteit

$$\mu_k^{i,j} = N_k^{i,j} - N_k^{i-1,j}$$

heeft, voor $j \neq \infty$ en

$$\mu_k^{i,\infty} = \beta_k^{i,\infty} - \beta_k^{i-1,\infty}.$$

Een persistentiediagram geeft ons de mogelijkheid om de persistentie-intervallen op een overzichtelijke manier weer te geven in een figuur. Hierbij plotten we een persistentie-interval $(i, j) \in \text{Dgm}_k(\mathcal{F})$, met $j < \infty$ als een punt in $\mathbb{R}^2$, en voegen we persistentie-intervallen van de vorm $(i, \infty) \in \text{Dgm}_k(\mathcal{F})$ toe als rode 'ruitjes', aan de top van het diagram, boven i. Zie figuur 2 voor de grafische weergave van het persistentiediagram, behorende bij een VR-filtratie uit onze toepassing.
Intuïtief geeft $N_k^{i,j}$ het aantal voortbrengende klassen van H_k^i aan dat in of voor Δ^i geboren zijn, en nog steeds leven in Δ^j. Elke multipliciteit $\mu_k^{i,j}$ geeft op die manier het aantal voortbrengende klassen aan dat in Δ^i geboren zijn, en nog niet uitgestorven zijn in Δ^j. In de volgende sectie zullen we deze intuïtieve eigenschappen formaliseren en bewijzen, waaruit ook zal volgen dat een persistentie-diagram onafhankelijk is van de gebruikte compatibele ordening.

3.3 Eigenschappen van persistentie-diagrammen

Vanaf nu nemen we alle homologieën over een **lichaam** K. Het voordeel is dat homologieën nu vectorruimtes zijn, en daarmee volledig vastgelegd zijn door hun dimensie, die simpelweg de al eerder genoemde Betti-getallen zijn. Omdat we met genestelde simpliciale complexen werken, is het logisch om ook de bases hierbij aan te laten sluiten.

Definitie 3.17 (Compatibele basis)**.** Laat $\mathcal{F}$ een filtratie zijn en $k \in \mathbb{N}$. Een *compatibele k-basis* $\mathcal{B}_k$ van F is een collectie bases $(B_k^i)_{i\in\mathbb{N}}$ voor respectievelijk de homologieën H_k^i, zodat voor elke $i \in \mathbb{N}$ geldt $f^{i,i+1}(B_k^i) \subseteq B_k^{i+1}$.

Merk op dat we altijd een compatibele k-basis (voor elke dimensie k) voor een filtratie kunnen construeren, door voor $i = 0$ een basis B_k^0 te kiezen voor H_k^0, en vervolgens voor elke $i > 0$ de bekende basisuitbreidingsstelling uit de lineaire algebra toe te passen op $f_k^{i-1,i}(B_k^{i-1})$. De intuïtie van de $N_k^{i,j}$ laat zich nu als volgt beschrijven.

Lemma 3.18. *Zij $\mathcal{F}$ een filtratie met compatibele ordening* ord*, $k \in \mathbb{N}$ en $\mathcal{B}_k$ een compatibele k-basis. Voor $0 \leq i < j < \infty$ geldt*

$$N_k^{i,j} = \sum_{i' \leq i} \#\{b \in B_k^{i'} : b \text{ geboren in } \Delta^{i'} \text{ en sterft uit in } \Delta^j\}.$$

Bewijs. Merk op dat het verschil in Betti-getallen $\beta_k^{i,j-1} - \beta_k^{i,j}$ het verschil in dimensie aangeeft van de persistente homologieën $H_k^{i,j-1}$ en $H_k^{i,j}$. Dit verschil is gelijk aan het verschil in kardinaliteit van $f^{i,j-1}(B_k^i) \subseteq B_k^{j-1}$ en $f_k^{i,j}(B_k^i) \subseteq B_k^j$. Dit laatste verschil kan alleen ontstaan als klassen uit $f^{i,j-1}(B_k^i)$ samen gaan met andere klassen in $f_k^{i,j}(B_k^i)$, wat precies overeen komt het het uitsterven van die klassen, omdat de compatibiliteit van de basis $\mathcal{B}_k$ ervoor zorgt dat we altijd de oudere klasse ook in de basis hebben zitten. Daarnaast is het duidelijk dat de uitstervende klassen geboren zijn in een zekere $\Delta^{i'}$, met $0 \leq i' \leq i$. □

Gevolg 3.19. *Voor een filtratie $\mathcal{F}$ met compatibele ordening* ord*, $k \in \mathbb{N}$ een compatibele basis $\mathcal{B}_k$ geldt voor alle $0 \leq i < j < \infty$*

$$\mu_k^{i,j} = \#\{b \in B_k^i : b \text{ geboren in } \Delta^i \text{ en sterft uit in } \Delta^j\}.$$

Bewijs. Voor $i = 0$ volgt dit direct uit de definitie van $N_k^{i,j}$ en lemma 3.18. Voor $i > 0$ volgt dit ook uit lemma 3.18, door het verschil van de sommen te nemen. □

Lemma 3.20. *Voor een filtratie $\mathcal{F}$ met compatibele ordening* ord*, $k \in \mathbb{N}$ een compatibele basis $\mathcal{B}_k$ geldt voor alle $i \in \mathbb{N}$*

$$\mu_k^{i,\infty} = \#\{b \in B_k^i : b \text{ geboren in } \Delta^i \text{ en sterft nooit uit}\}.$$

Bewijs. Merk op dat we analoog aan het bewijs van lemma 3.18 hebben dat $\beta_k^{i,\infty}$ het aantal verschillende basiselementen van B_k^i is dat in elke B_k^j, $j > i$ voorkomt, en dus nooit uitsterven. Deze klassen zijn elk geboren in een $\Delta^{i'}$, met $i' \leq i$, en nemen we het verschil zoals in de definitie van $\mu_k^{i,\infty}$, dan krijgen we precies bovenstaande uitspraak. □

Aan de hand van de definitie van de $\mu_k^{i,j}$ zagen we al dat deze onafhankelijk waren van de compatibele ordening op $\mathcal{F}$. Met gevolg 3.19 en lemma 3.20 krijgen we nu ook een karakterisatie aan de hand van het persistentie-intervallen van basiselementen uit een compatibele basis.

Lemma 3.21 (Hoofdlemma van persistente homologie)**.** *Zij $\mathcal{F}$ een filtratie, $k \in \mathbb{N}$ en $0 \leq m \leq n$. Dan geldt:*

$$\beta^{m,n} = \sum_{i \leq m} \sum_{j > n} \mu_k^{i,j}.$$

Bewijs. Laat $\mathcal{B}_k$ een compatibele basis van $\mathcal{F}$ zijn. We vinden de volgende serie van gelijkheden:

$$\begin{aligned}
\beta_k^{m,n} &:= \text{rang}(\text{im}(f_k^{m,n})) \\
&\overset{(i)}{=} \#f_k^{m,n}(B_k^m) \\
&\overset{(ii)}{=} \#\{b \in B_k^m : \text{sterfte}(b) > n\} \\
&\overset{(iii)}{=} \#\{b \in B_k^i \text{ geboren in } \Delta^i, \text{ met } i \leq m \text{ en sterfte}(b) > m\} \\
&\overset{(iv)}{=} \sum_{i \leq m} \sum_{j > n} \mu_k^{i,j}
\end{aligned}$$

Hierbij is (i) simpelweg een herschrijving in termen van de basiselementen. Stap (ii) maakt gebruik van de definitie van sterfte, en stap (iii) combineert geboorte met de compatibele basis. Tot slot gebruiken we in stap (iv) gevolg 3.19 en lemma 3.20. □

Lemma 3.21 vertelt ons dat voor een filtratie, waarvan de homologieën over een lichaam worden genomen, de volledige persistente homologie (gekarakteriseerd door de Betti-getallen) vastligt in het persistentiediagram. In het volgende hoofdstuk zullen we dan ook een algoritme beschrijven die de persistentie-intervallen van een compatibele basis, en daarmee het persistentiediagram, kan berekenen.

4 Berekenen van persistente homologie

We gaan nu een algoritme beschrijven dat gebruikt wordt voor het berekenen van persistente homologie. Vanaf nu zullen we werken over het lichaam $\mathbb{F}_2$ en beschouwen we enkel nog eindige filtraties.

4.1 Algoritme

Laat $\mathcal{F}$ een eindige filtratie zijn met compatibele ordening ord. In het bijzonder heeft $\Omega(\mathcal{F})$ nu een eindig aantal elementen n, die we nummeren als $\omega_1, ..., \omega_n$, zodanig dat voor $i, j \in \{1, ..., n\}$ geldt $i < j \Leftrightarrow \text{ord}(\omega_i) < \text{ord}(\omega_j)$. Het algoritme is gebaseerd op een matrixreductie, dus we voeren eerst de volgende notatie in: voor een matrix M schrijven we M^i voor de i-de rij, M_j voor de j-de kolom, en M_j^i voor de (i,j)-de component.
Nu kijken we naar de $n \times n$-matrix $\partial(\mathcal{F})$, of kortweg ∂, diens elementen gegeven worden door:

$$\partial_j^i = \begin{cases} 1, & \text{als } \omega_i \text{ een zijde is van } \omega_j \text{ van codimensie 1;} \\ 0, & \text{anders.} \end{cases}$$

De rijen van ∂ geven dus precies de randen van de σ_i, en door de compatibele ordening weten we dat ∂ een bovendriehoeksmatrix is. Merk op dat ∂ de matrix is behorende bij de volgende lineaire afbeelding:

$$\begin{aligned}
\partial \colon \bigoplus_{k \in \mathbb{N}} C_k &\to \bigoplus_{k \in \mathbb{N}} C_k, \\
\bigoplus_{k \in \mathbb{N}} \sigma_k &\mapsto \bigoplus_{k \in \mathbb{N}} \partial_k(\sigma_k).
\end{aligned}$$

Door kolomoperatoren uit te voeren, reduceren we ∂ tot een matrix R, waaraan we de persistentie-intervallen van een compatibele basis kunnen aflezen. Eerst volgt nog een belangrijke definitie voor het algoritme, die ook van belang zal zijn bij de interpretatie van het resultaat ervan.

Definitie 4.1 (low_M)**.** Zij M een $m \times n$-matrix. We definiëren de functie $\text{low}_M : \{1, ..., n\} \to \{0, ..., m\}$ als volgt:

$$\text{low}_M(j) := \begin{cases} 0, & \text{als } M_j \text{ alleen nullen heeft} \\ \max\{i \in \{1, 2, ..., m\} : M_j^i \neq 0\}, & \text{anders.} \end{cases}$$

We hebben nu alle benodigde notatie ingevoerd, en kunnen beschrijven hoe de reductie verloopt (zie ook [4])

Algorithm 1 Reductie-algoritme

```
R ← ∂
for j ← 1 to n do
    while there is k ∈ {1, ..., j} with 0 ≠ low_R(k) = low_R(j) do
        R_j ← R_j + R_k
    end while
end for
return R
```

Aangezien er voor elke kolom ∂_j er niet meer dan $j-1$ kolommen van R bij op worden geteld in het algoritme, en het aantal kolommen eindig is, volgt direct dat het reductie-algoritme termineert.

4.2 Interpretatie van de gereduceerde matrix

Passen we het algoritme toe op een filtratie $(\mathcal{F}, \text{ord})$ met compatibele ordening, en houden we in elke stap bij welke kolommen van ∂ uiteindelijk optellen tot een kolom van R, dan kunnen we deze informatie als volgt opslaan in een $n \times n$ matrix V: we hebben $V_j^i = 1$ als de kolom ∂_i (indirect) via het algoritme bij kolom R_j wordt opgeteld, en $V_j^i = 0$ anders. De berekening van R is dan verkort op te schrijven als:

$$R = \partial \cdot V.$$

De kolommen van R geven ons informatie over de cycli en grenzen die ontstaan bij het stuk voor stuk toevoegen van de σ_i in de geordende filtratie $\hat{F}$. Laat $j \in \{1, ..., n\}$ zijn, met $\omega_j \in \Omega$ een k-simplex. We onderscheiden twee mogelijkheden:

(1) Er geldt $\text{low}_R(j) = 0$. In dit geval geldt $\partial V_j = 0$, dus $V_j = \bigoplus_{n \in \mathbb{N}} \sigma_n$ is een directe som van cycli. Specifiek weten we dat $V_j^j = 1$ en aangezien ω_j een k-simplex is, volgt dat σ_k een k-cyclus is. Deze cyclus bestond nog niet in $\tilde{\Delta}^{j-1}$, dus σ_k is geboren in $\tilde{\Delta}^j$.

(2) Er geldt $\text{low}_R(j) = i \neq 0$. Schrijven we wederom $V_j = \bigoplus_{n \in \mathbb{N}} \sigma_n$, dan weten we nu dat geldt $\tau := \partial_k \sigma_k$ een niet-lege k-grens is (er geldt namelijk $\tau_i = 1$, want $R_j^i = 1$). Het toevoegen van ω_j aan $\tilde{\mathcal{F}}$ zorgt er dus voor dat τ uitsterft in $\tilde{\Delta}^j$. Daarnaast weten we ook dat de klasse die uitsterft, geboren was in $\tilde{\Delta}^i$, aangezien ω_i daarvoor nog niet in filtratie zat.

We kunnen dus perisitentie-intervallen aflezen van $\tilde{\mathcal{F}}$ uit de gereduceerde matrix R. Specifiek hebben we een persistentie-interval (i, j), met $j \neq \infty$ als geldt $\text{low}_R(j) = i$. Daarnaast krijgen we ook persistentie-intervallen van de vorm (i, ∞) als $\text{low}_R(i) = 0$ en $i \neq \text{low}_R(j)$, voor alle $j \in \{1, ..., n\}$. Met de methode die genoemd is in het vorige hoofdstuk, kunnen we de persitentie-intervallen van de oorspronkelijke filtratie $\mathcal{F}$ bepalen, door te kijken naar de indices $\text{in}(\omega_i)$ en $\text{in}(\omega_j)$

Deel II
Toepassing

5 Methode

Het is nu tijd om de mogelijkheden van persistente homologie verder te ontdekken. We lichten kort toe wat voor data we gebruiken en op welke manier het experiment wordt opgezet.

5.1 Data

We hebben drie verschillende soorten data gebruikt voor ons onderzoek. Alle drie bestaan ze uit "point cloud data" (PCD), als eindige deelverzameling van de $\mathbb{R}^3$. We lichten kort toe hoe we aan de data zijn gekomen, en welke restricties op de datasets hebben gedaan.

Gm Voor onze echte data hebben we een deel van de GAMA (Galaxy And Mass Assembly [3]) catalogus gebruikt. Specifiek gebruiken we de metingen van de G09, G12 en G15 regio's. Van elk geobserveerd object weten we onder andere de rechte klimming (RA), de declinatie (DEC), de roodverschuiving (z) en de absolute magnitude (M). Deze data heeft twee complicaties: op kleine roodverschuiving is de dichtheid te hoog, en op grote waarden voor z nemen we niet meer alle sterrenstelsels waar. Om hier voor te corrigeren, nemen we zoals in [10] alleen de stelsels waarvoor geldt $-21.8 < M < -20.1$ en $0.039 < z < 0.263$. Op die manier krijgen we een representatieve populatie. De data van de drie losse stukken geven we aan met Gm[0], Gm[1] en Gm[2]. Een projectie van hoe de data eruit ziet is te zien in figuur 1. Na de magnitude- en roodverschuivingsbeperkingen, zijn zowel het aantal stelsels, als het rechteklimming- en declinatiebereik per regio weergegeven in tabel 1.

Tabel 1: Parameters GAMA-data

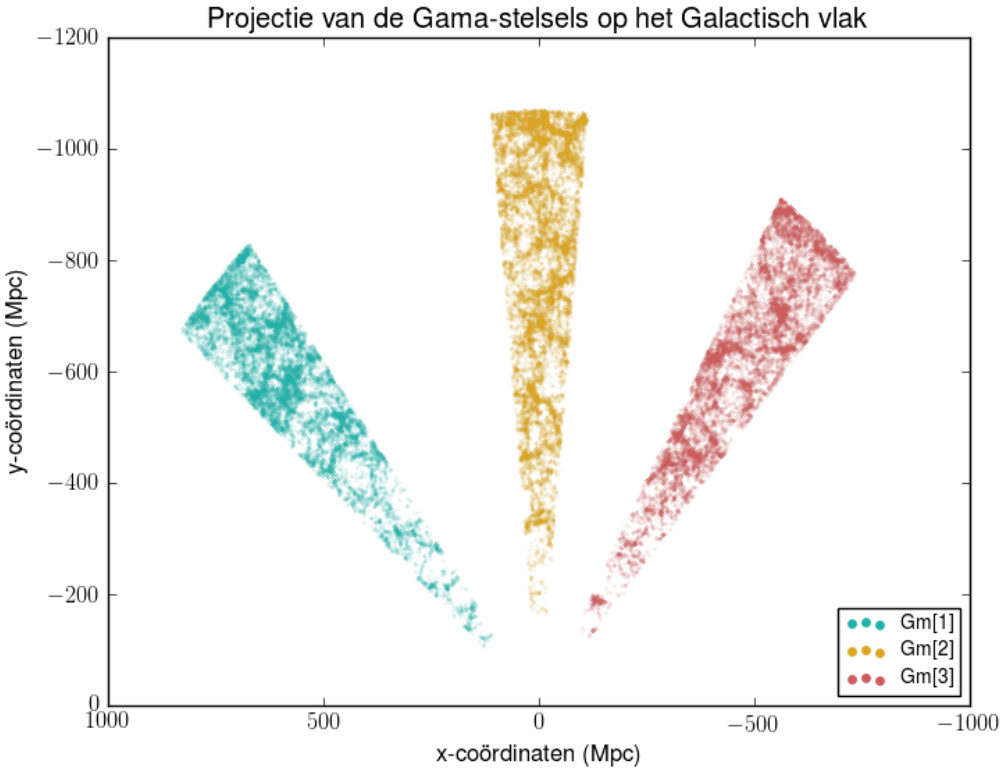
	Gm[0]	Gm[1]	Gm[2]
Aantal stelsels	14723	16313	15264
$RA(^\circ)$	129.0 - 141.0	174.0 - 186.0	211.5 - 223.5
$DEC(^\circ)$	-2 - $+3$	-3 - $+2$	-2 - $+3$

Gegevens van de drie verschillende GAMA regio's, respectievelijk G09, G12 en G15. Daarnaast zijn er nog beperkingen op de absolute magnitude en roodverschuiving geplaatst ([10])

Ml De gesimuleerde data die we willen vergelijken met de GAMA-data, hebben we van de milli-Millennium simulation ([2], [9]). In een kubusvormige box met zijde 62,5 Megaparsec worden simulaties gedaan met 270^3 punten. Via een sql-query zijn daar alle ontstane sterrenstelsels met dezelfde helderheidsrestrictie als voor de GAMA-data uitgehaald. De simulatie werkt met verschillende snapshots: in 64 tijdstappen zijn de tussenresultaten van de simulatie opgeslagen. We gebruiken de snapshots 25, 50 en 63 om een idee te krijgen voor de verschillen in de evolutie over de tijd, en noteren deze respectievelijk als Ml[25], Ml[50] en Ml[63].

Rd Ter controle willen we ook uniform verdeelde willekeurige data vergelijken met de GAMA-data. Hiervoor verdelen we $n = \#\text{Gm}[0]$ punten uniform over een gebied van dezelfde

Figuur 1: Projectie van de stelsels van de verschillende GAMA regio's



Van de drie gebruikte regio's, respectievelijk G09, G12 en G15, zijn de waargenomen stelsels uit de GAMA-catalogus geprojecteerd op het Galactisch vlak weergegeven. Het betreft hier de stelsels die binnen absolute magnitude range van $-21.8 < M < -20.1$ en een roodverschuivingsbereik hebben van $0.039 < z < 0.263$. De coördinaten zijn hierbij omgezet naar Euclidische coördinaten, zodat deze direct compatibel zijn met Perseus.

afmetingen als die van Gm[0]. Hier hebben we 100 samples van gemaakt. We zullen deze sets aangeven met Rd[i], met $i \in \{0, ..., 99\}$.

5.2 Analyse van de data

Van elk van bovenstaande datasets gaan we een VR-filtratie maken. Vervolgens passen we het reductie-algoritme toe om de persistentie-intervallen te vinden. Voor elke VR-filtratie nemen we $\varepsilon_i = i/100$, voor alle $i \in \{0, ..., 200\}$, en $\varepsilon_i = 2$ voor alle $i > 200$. Op deze manier krijgen we een eindige VR-filtratie, waarbij we tot een maximale koppellengte van 2 Megaparsec gaan.
Het construeren van de VR-filtratie en de berekening van de persistente homologie doen we aan de hand van het software-pakket *Perseus* [12]. Daarnaast voert Perseus ook *Morse-theoretische reducties* uit om de berekening te versnellen. De werking hiervan valt echter buiten het bereik van dit onderzoek. Voor de theorie erachter kan gekeken worden naar [11]. Aangezien we de ε_i steeds met dezelfde stapgrootte vergroten, is het voldoende om naar de *index*-persistentie te kijken, zoals we die in 3.15 hebben gedefinieerd. Omdat we voornamelijk geïnteresseerd zijn naar de persistentie van homologieklassen, voeren we een KS-test uit, met als steekproeven de lengtes van de persistentie-intervallen van de verschillende datasets. Hierbij beperken we ons tot de eindige persistenties, aangezien de oneindige persistenties simpelweg nog niet zijn gestorven in $\Delta_2(S)$, met S een dataset, maar er kan niet achterhaald worden of er een koppellengte is vanaf wanneer dit wel het geval is. We passen de KS-test toe voor de dimensies 0, 1 en 2. We vergelijken ook de verschillen tussen de resultaten voor de verschillende dimensies.

Notatie 5.1. We noteren len(Gm[i]) voor de verzameling van eindige lengtes van persistentie-intervallen uit $\mathrm{Dgm}_k(\mathcal{F})$ van de VR-filtratie $\mathcal{F} := \mathcal{F}_{\mathrm{VR}}(\mathrm{Gm}[i])$. Analoog schrijven we len(Ml[i]) en len(Rd[i]). Daarnaast schrijven we $\mathrm{len}(\mathrm{Gm}) := \bigcup_i \mathrm{len}(\mathrm{Gm}[i])$, en analoog definiëren we len(Rd).

Voor de lezer die niet bekend is met de KS-test, wijden we hier in de volgende paragraaf op uit.

5.3 Kolmogorov-Smirnovtest

De Kolmogorov-Smirnovtest, of kortweg KS-test, is een statistische toets die gebruikt wordt om te kijken of twee gegeven steekproeven uit dezelfde verdeling getrokken kunnen zijn (specifiek gebruiken we dus de "2-sample KS-test", zie ook [6]). Laat $\mathcal{X} = \{X_1, ..., X_n\}$ een onafhankelijke en gelijk verdeelde steekproef zijn van een (onbekende) verdeling $F_{\mathcal{X}}$. Een dergelijke steekproef noemen we ook een *aselecte steekproef* (van $F_{\mathcal{X}}$). De *empirische verdelingsfunctie* $\hat{F}_{\mathcal{X}}$ van de steekproef $\mathcal{X}$ wordt gegeven door:

$$\hat{F}_{\mathcal{X}}(x) = \frac{1}{n} \sum_{i=1}^{n} I_x(X_i),$$

waarbij I een indicatorfunctie is:

$$I_x(X_i) = \begin{cases} 1, & \text{als } X_i \leq x \\ 0, & \text{anders} \end{cases}$$

Hebben we nu twee aselecte steekproeven $\mathcal{X}$ en $\mathcal{Y}$, dan willen we de volgende nulhypothese, H_0, testen tegen $H1$:

$$H0 : F_{\mathcal{X}} = F_{\mathcal{Y}}$$
$$H1 : F_{\mathcal{X}} \neq F_{\mathcal{Y}}.$$

Hiervoor gebruiken we de *Kolmogorov-Smirnov*-statistiek (KS-statistiek) $M_{\mathrm{KS}}(\mathcal{X}, \mathcal{Y})$, die gegeven wordt door het supremum van het verschil tussen de empirische verdelingsfuncties van $\mathcal{X}$ en $\mathcal{Y}$:

$$M_{\mathrm{KS}}(\mathcal{X}, \mathcal{Y}) = \sup_x \left| \hat{F}_{\mathcal{X}}(x) - \hat{F}_{\mathcal{Y}}(x) \right|$$

Schrijf $m = |\mathcal{X}|$ en $n = |\mathcal{Y}|$. Voor grote waarden van m en n krijgen we de volgende kritieke waarde ([16]):

$$D_\alpha(m, n) = c(\alpha)\sqrt{\frac{m+n}{mn}},$$

waarbij we de nulhypothese verwerpen als $M_{\mathrm{KS}}(\mathcal{X}, \mathcal{Y}) > D_\alpha(m, n)$. De waarden van $c(\alpha)$ worden gegeven in tabel 2, en zijn onafhankelijk van $F_{\mathcal{X}}$ en $F_{\mathcal{Y}}$. Voor waarden van m en n tot en met 12, is bovenstaande asymptotische waarde van $D_\alpha(m, n)$ niet toereikend, en kunnen de waarden in een tabel worden opgezocht ([16]).

Tabel 2: Waarden van $c(\alpha)$

α	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

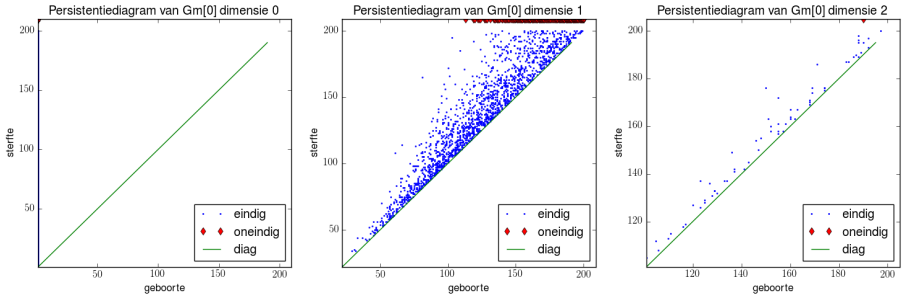
6 Resultaten

We hebben drie keer een aantal KS-tests uitgevoerd. De eerste test dient als controle, waarin we de verschillende Gm[i] met elkaar vergelijken. Vervolgens willen we weten of er ook een verschil te zien is tussen Gm en data waarvan we de verdeling weten; Rd. Tot slot vergelijken we Gm met de verschillende snapshots van Ml om te zien hoe goed de simulatie op verschillende tijdstippen overeenkomen met echte data.

6.1 Controleproef I: Gm tegen Gm

Paarsgewijs hebben we de verschillende Gm[i] met elkaar vergeleken. Het verwachte resultaat van de KS-test is dat de nulhypothese niet verworpen wordt, aangezien de structuur van het kosmisch web op zekere schaal als isotroop wordt beschouwd. De persistentiediagrammen van Gm[0] zijn weergegeven in figuur 2, om een idee te krijgen hoe het persistentiediagram eruit ziet. Zoals verwacht van het VR-complex, zijn alle 0-de klassen geboren in Δ_0(Gm[0]). Bij de hogere dimensies is meer spreiding te zien.
De empirische verdelingsfuncties van de len(Gm[i]) zijn per dimensie weergegeven in figuur 3. In tabel 3 staan de resultaten van de verschillende KS-testen. Zoals te zien is wordt geen enkele nulhypothese verworpen voor $\alpha \leq 0.01$. Alleen bij Gm[0] tegen Gm[2] wordt de nulhypothese voor de 0-dimensionale persistentie-intervallen verworpen bij een α-waarde van 0.025. We zien dat de verschillende empirische verdelingen sterk op elkaar lijken voor een groot deel van de data. De afwijking bij de 0-dimensionale klassen wordt waarschijnlijk veroorzaakt door "kosmische variantie" (zie [13]). De kosmische variantie is de onzekerheid in de waargenomen lokale dichtheidsverdeling, veroorzaakt door dichtheidsfluctaties op grote schaal. Hierdoor zijn de 3 regio's niet geheel representatief voor de totale dichtheidsverdeling, en kunnen relatief grote verschillen bestaan. Om hier (deels) voor te compenseren vergelijken we in de volgende tests niet met een enkele regio van de GAMA-data, maar gebruiken we de volledige verdeling len(Gm) van lengtes van persistentie-intervallen.

Figuur 2: Persistentie diagrammen Gm[0]



Persistentiediagram, behorende bij de VR-filtratie met als puntenverzameling de stelsels uit Gm[0]. In het eerste diagram staan de 0-de persistentie-intervallen van de basiselementen. In het tweede en derde diagram staan analoog de 1- en 2-dimensionale persistentie-intervallen respectievelijk.

Tabel 3: Resultaten KS-test van de verschillende len(Gm[i]) tegen elkaar

A: len(Gm[0]) tegen len(Gm[1])

$k\backslash\alpha$	0.10	0.05	0.025	0.01	0.005	0.001
0	×	×	✓	✓	✓	✓
1	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓

B: len(Gm[0]) tegen len(Gm[2])

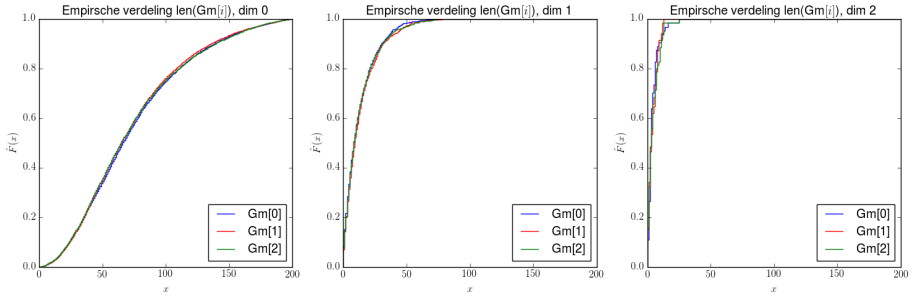
$k\backslash\alpha$	0.10	0.05	0.025	0.01	0.005	0.001
0	×	×	×	✓	✓	✓
1	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓

C: len(Gm[1]) tegen len(Gm[2])

$k\backslash\alpha$	0.10	0.05	0.025	0.01	0.005	0.001
0	✓	✓	✓	✓	✓	✓
1	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓

Per tabel zijn twee verschillende len(Gm[i]) tegen elkaar getoetst met de KS-test. Een vinkje induceert dat de nulhypothese onder die α-waarde wordt geaccepteerd. Een kruis duidt op verwerping van de nulhypothese.

Figuur 3: Empirische verdelingsfuncties len(Gm)



De empirische verdelingsfuncties van elke len(Gm[i]), voor de dimensies 0, 1 en 2.

6.2 **Controlepoef II:** Gm tegen Rd

Nu we weten wat de resultaten zijn van de KS-tests van de verschillende len(Gm[i]), willen we deze vergelijken met een verdeling waarvan we weten dat deze anders is. Het verwachte resultaat is dat de nulhypothese nu verworpen wordt op hogere significantie. Omdat we weten dat de verschillende Rd[i] uit dezelfde verdeling komen, testen we direct de nulhypothese $H1 : F_{\text{len(Gm)}} = F_{\text{len(Rd)}}$ tegen het alternatief $H1 : F_{\text{len(Gm)}} \neq F_{\text{len(Rd)}}$. De resultaten van de KS-tests zijn te zien in tabel 4, en de empirische verdelingen zijn geplot in figuur 5. Ook hebben we nog de persistentiediagrammen van Rd[0] weergegeven in figuur 4, om op empirische gronden de willekeurige data met de GAMA-data te kunnen vergelijken.
Bij het vergelijken van figuur 2 met figuur 4 valt direct het lage aantal dimensie-2-persistentie-intervallen van Rd[0] op. Dit is geen toeval; het gemiddelde aantal van deze klassen ligt op 4.5 ± 2.3 per dataset Rd[i], met een hoogst voorkomende waarde van 11. Dat wijkt zeer significant af van het gemiddelde aantal dimensie-2-persistentie-intervallen van de Gm[i]: die ligt namelijk op 68 ± 3. In de 1-ste persistentie-intervallen is ook lichtelijk een verschil in spreiding te zien, maar dit is moeilijk met zekerheid te zeggen, dus daarvoor kijken we naar de KS-test. Evenals voor de dimensie-1-intervallen; die lijken vrijwel identiek in het persistentiediagram.

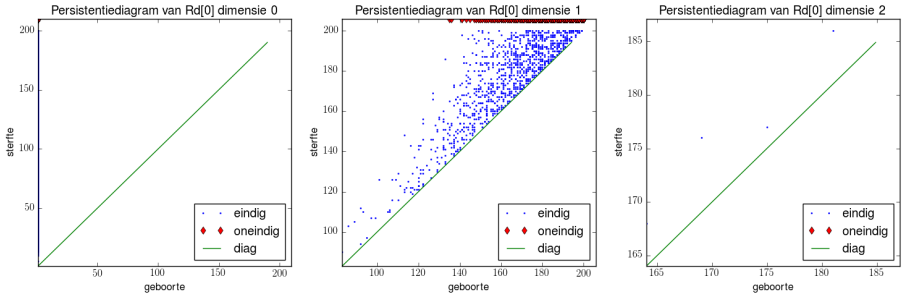
Tabel 4: Resultaten KS-test op len(Gm) tegen len(Rd)

$k \backslash \alpha$	0.10	0.05	0.025	0.01	0.005	0.001
0	×	×	×	×	×	×
1	×	×	×	×	×	×
2	✓	✓	✓	✓	✓	✓

Hier zijn de resultaten van de KS-tests van de verschillende dimensies van len(Gm) tegen len(Rd). Een kruis indiceert een verwerping van de nulhypothese $F_{\text{len(Gm)}} = F_{\text{len(Rd)}}$. Een vink indiceert het accepteren van de nulhypothese.

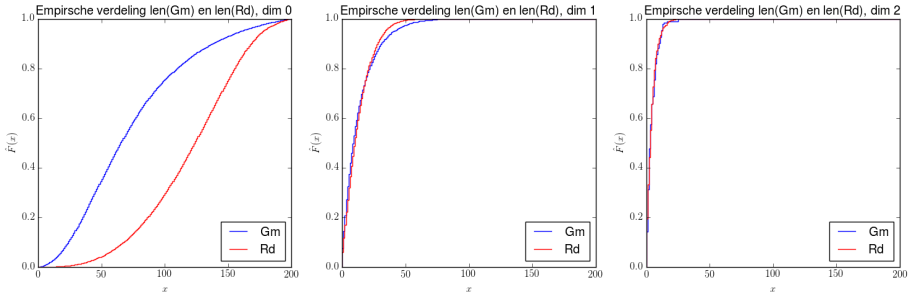
De resultaten van de KS-tests (tabel 4) komen goed overeen met de verwachte uitkomsten, als we kijken naar plots van de empirische verdelingen (zie figuur 5). Enigszins verrassend zijn de resultaten precies het complement van wat we in eerste instantie empirisch aan de hand van de

Figuur 4: Persistentie diagrammen Rd[0]



Persistentiediagram, behorende bij de VR-filtratie met als puntenverzameling de stelsels uit Rd[0]. In het eerste diagram staan de 0-de persistentie-intervallen van de basiselementen. In het tweede en derde diagram staan analoog de 1- en 2-dimensionale persistentie-intervallen respectievelijk. Dit is slechts één van de datasets van Rd, en wordt niet op zichzelf toegepast in een KS-test, maar is bedoeld voor empirische vergelijking met Gm[0].

Figuur 5: Empirische verdelingsfuncties van len(Gm), en len(Rd)



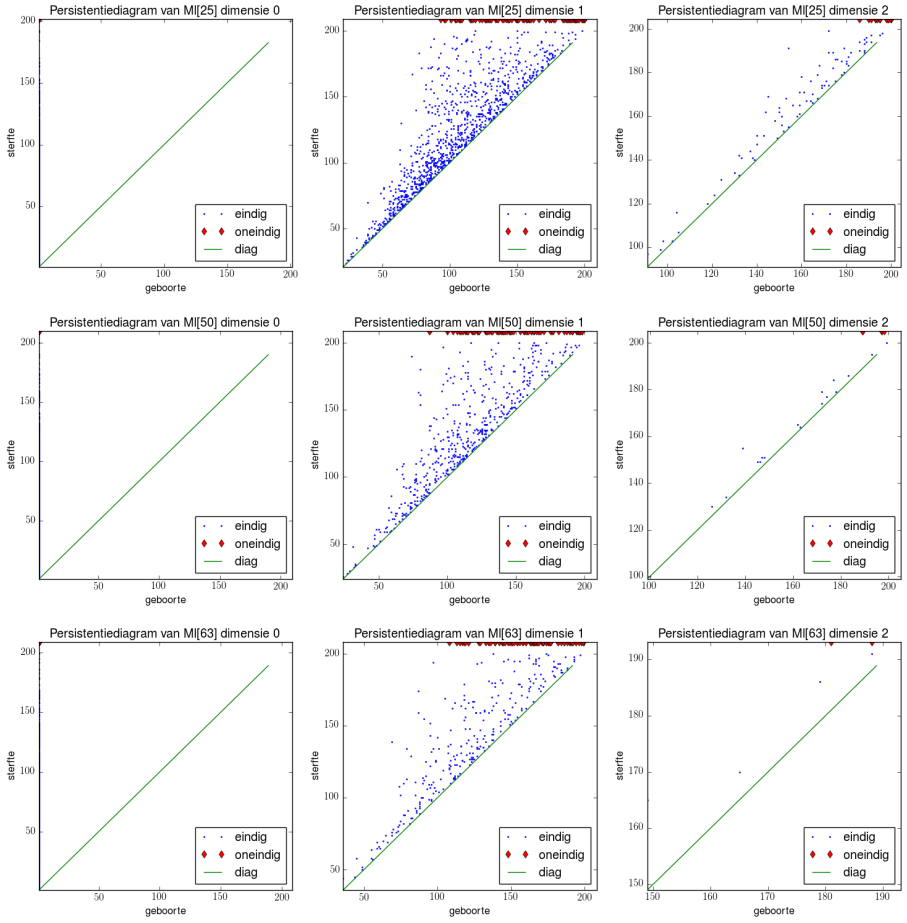
De empirische verdelingsfuncties van len(Gm) en len(Rd), in de dimensies 0, 1 en 2.

persistentiediagrammen konden afleiden. We zien dus dat de KS-test een goede aanvulling is bij het vergelijken van persistentiediagrammen; het levert het verwachte resultaat bij een test waarvan van te voren bekend was dat de verdelingen anders waren.

6.3 De test: Gm tegen Ml

We doen nu 3 tests: we vergelijken len(Gm) met len(Ml[i]), voor $i \in \{25, 50, 63\}$. Hier hebben we voor elke snapshot de persistentiediagrammen gemaakt, en deze zijn te zien in figuur 6.

Figuur 6: Persistentiediagrammen Ml[25], Ml[50] en Ml[63]



De persistentiediagrammen van dimensies 0, 1 en 2 (van links naar rechts), van de VR-filtraties milli-Millennium snapshots 25, 50 en 63 (van boven naar beneden).

Ook hiervan hebben we de verdelingsfuncties geplot, zie figuur 7. De toetsresultaten laten ons vermoeden dat door de evolutie in tijd, de milli-Millennium-data sterker gaat lijken op de GAMA-data. Dit wordt op het zicht bevestigd in de figuur 7, waarin te zien is dat de verdelingsfunctie van len(Ml[63]), die bij de laatste snapshot 63 hoort, het dichtst bij die van len(Gm) ligt. We

Tabel 5: Resultaten KS-tests van len(Gm) tegen respectievelijk len(Ml[25, 50, 63])

A: len(Gm) tegen len(Ml[25])

$k\backslash\alpha$	0.10	0.05	0.025	0.01	0.005	0.001
0	×	×	×	×	×	×
1	×	×	×	×	×	×
2	×	×	✓	✓	✓	✓

B: len(Gm) tegen len(Ml[50])

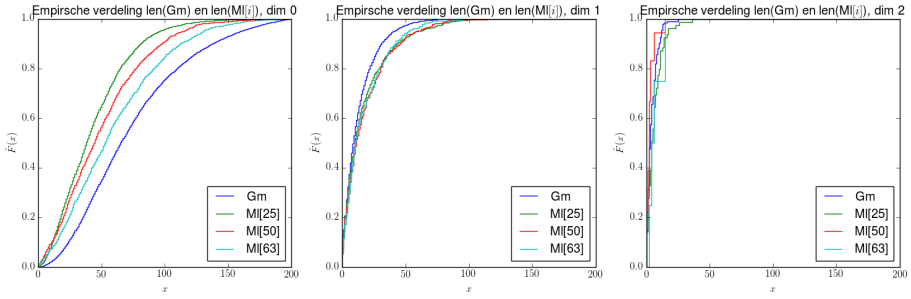
$k\backslash\alpha$	0.10	0.05	0.025	0.01	0.005	0.001
0	×	×	×	×	×	×
1	×	×	×	×	×	×
2	✓	✓	✓	✓	✓	✓

C: len(Gm[) tegen len(Ml[63])

$k\backslash\alpha$	0.10	0.05	0.025	0.01	0.005	0.001
0	×	×	×	×	×	×
1	×	×	×	×	×	✓
2	✓	✓	✓	✓	✓	✓

De resultaten van de KS-tests van len(Gm) tegen zowel len($Ml[i]$), voor snapshots $i \in \{25, 50, 63\}$. Een kruis betekent dat de nulhypothese wordt verworpen met bovenstaande α, en analoog bij wordt deze bij een vink geaccepteerd.

Figuur 7: Empirische verdelingen len(Gm) en len(Ml[i])



De empirische verdelingen van len(Gm) en len(Ml[i]), voor alle $i \in \{25, 50, 63\}$.

hebben echter genoeg reden om de nulhypothese $F_{\text{len(Gm)}} = F_{\text{len(Ml}[i])}$ voor elke i te verwerpen.

7 Discussie

In de eerste controleproef hebben we gezien dat de KS-test levert dat twee verschillende stukken van de GAMA-data bijna altijd tot dezelfde distributie van persistentie-intervallen leiden, ondanks dat de stukken van verschillende grootte en vorm waren. Er was een enkele uitzondering, maar deze werd waarschijnlijk veroorzaakt door de kosmische variantie. In de tweede controleproef zien we een sterk verschil tussen de persistente homologie van de GAMA-data en de uniform verdeelde willekeurige data. Dit laatste is wat we verwachten dat de GAMA-data zou zijn, mochten er in het geheel geen interacties als zwaartekracht tussen de sterrenstelsels van het kosmisch web plaatsvinden. Uiteraard is er wel interactie tussen de stelsels, en verwachtten we dat KS-test de nulhypothese $F_{\text{len(Gm)}} = F_{\text{len(Rd)}}$ zou verwerpen. Het resultaat is in overeenstemming met onze verwachting.
Beide controleproeven versterken ons vertrouwen in de KS-test op verzamelingen van lengtes van persistentie-intervallen. Bij het vergelijken tussen de geobserveerde GAMA-data met de gesimuleerde milli-Millennium-data, zien we nu een significant verschil tussen de empirische verdelingsfuncties. Als conclusie kunnen we hieruit trekken dat de gesimuleerde data niet volledig overeenkomt in topologische eigenschappen als geobserveerde data. Wel zagen we dat de verschillen kleiner werden wanneer we naar een later stadium van de simulatie keken. Dit is echter niet de enige oplossing die tot een beter model zou kunnen leiden. We zien namelijk ook in figuur 6 dat het aantal dimensie-2-persistentie-intervallen sterk is gedaald naarmate we verder in de snapshots kwamen. De reden dat onze test de nulhypothese verwerpt, vindt waarschijnlijk zijn oorzaak in de kwaliteit van de simulaties. Er wordt nog veel onderzoek gedaan naar zogenaamde 'galaxy formation recipes' (zie [1]). Onder andere de verdeling van donkere materie en diens invloed op de vorming van sterrenstelsels wordt benaderd. Verder is het nog onduidelijk of de kwaliteit van de modellen verbeterd kan worden door alleen een betere modellering van astrofysische processen, of dat ook het kosmologisch model moet worden aangepast (zie [1] en [5]).
Verder zagen we bijna geen resultaten uit de 2-de persistente homologie-intervallen die de nulhypothese zouden verwerpen. Een mogelijke verklaring voor dit feit is dat de bijbehorende topologische eigenschappen, de voids, pas op grotere schaal dan 2 Megaparsec voorkomen en met de computationele beperkingen hier niet gedetecteerd konden worden. Wel konden we een significant verschil zien tussen de geobserveerde en de uniform verdeelde data in het aantal persistentie-intervallen dat per dataset ontstond.

Deel III
Slotbeschouwing

8 Samenvatting

We hebben de theorie van persistente homologie van de grond af opgebouwd aan de hand van filtraties van simpliciale complexen. We hebben daarbij enkele definities en lemma's veralgemeniseert en verbetered ten op zichte van het werk van Edelsbrunner [4]. We zagen dat de persistente homologieën volledig werden vastgelegd door het persistentiediagram, indien over een grondlichaam K wordt gewerkt. Dit leidt tot een eenvoudig $\mathcal{O}(n^3)$-algoritme, voor een eindige filtratie van Δ, waarbij $n = \#\Delta$.
Vervolgens hebben we het algoritme toegepast op kosmologische point cloud data, en hebben gezien dat de KS-test een geschikte test is om de verdelingen van lengtes van persistentie-intervallen te vergelijken. We concludeerden dat de test onderscheid kan maken tussen geobserveerde data van het kosmische web, afkomstig van de GAMA-catalogus en uniform verdeelde willekeurige data. Daarnaast hebben we ook de geobserveerde data vergeleken met gesimuleerde data van het milli-Millennium project, en zagen dat de huidige gesimuleerde modellen nog niet perfect zijn.

9 Toekomstperspectieven

Op dit moment worden de toepassingen van persistente homologie op grote datasets als GAMA nog sterk gehinderd door de computationele beperkingen van de bestaande software. Er zal meer onderzoek gedaan moeten worden naar efficiënte implementaties van het algoritme en optimalisaties op het gebied van geheugengebruik.
De theorie zelf heeft nu een goede karakterisatie, wanneer de homologieën over een grondlichaam K wordt genomen. Een natuurlijke vraag is hoe de persistente homologieën zich gedragen over andere grondringen. Daarnaast kan er gekeken worden naar Morse theorie [11] wat zou kunnen leiden tot meer inzichten in de eigenschappen van persistente homologie.
We hebben ook gezien dat de modelsimulaties van het kosmische web nog niet optimaal zijn. In de toekomst kunnen we de methode die we hier hebben toegepast, gebruiken om kosmologische modellen met elkaar, en met geobserveerde data, te vergelijken.

Referenties

[1] Alfonso Aragon-Salamanca, Carlos S. Frenk, Julio F. Navarro, and Stephen E. Zepf. A Recipe for Galaxy Formation Shaun Cole. (ii):1–37, 1994.

[2] Gabriella De Lucia and Jérémy Blaizot. The hierarchical formation of the brightest cluster galaxies. *Monthly Notices of the Royal Astronomical Society*, 375(1):2–4, 2007.

[3] Simon P. Driver, Peder Norberg, Ivan K. Baldry, Steven P. Bamford, Andrew M. Hopkins, Jochen Liske, Jon Loveday, and John A. Peacock. GAMA: Towards a physical understanding of galaxy formation. *Astronomy and Geophysics*, 50(5), 2009.

[4] Herbert Edelsbrunner and John Harer. *Computational topology.* 2008.

[5] George F R Ellis and Henk van Elst. Cosmological models. *NATO Adv. Study Inst. Ser. C. Math. Phys. Sci.*, 541:1–116, 1999.

[6] Eric D. Feigelson and Jogesh Babu. *Modern Statistical Methods for Astronomy*. 2012.

[7] Robert Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[8] Allen Hatcher. *Algebraic Topology*. 2001.

[9] Gerard Lemson. Halo and Galaxy Formation Histories from the Millennium Simulation : Public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the Λ CDM cosmogony. *arXiv preprint astro-ph/0608019*, 1:1–7, 2006.

[10] T. McNaught-Roberts, P. Norberg, C. Baugh, C. Lacey, J. Loveday, J. Peacock, I. Baldry, J. Bland-Hawthorn, S. Brough, S. P. Driver, a. S. G. Robotham, and J. a. Vazquez-Mata. Galaxy And Mass Assembly (GAMA): the dependence of the galaxy luminosity function on environment, redshift and colour. *Monthly Notices of the Royal Astronomical Society*, 445(2):2125–2145, 2014.

[11] Konstantin Mischaikow and Vidit Nanda. Morse Theory for Filtrations and Efficient Computation of Persistent Homology. *Discrete and Computational Geometry*, 50(2):330–353, 2013.

[12] Vidit Nanda. *Perseus, the Persistent Homology Software*, Aaccessed April 17, 2016.

[13] RS Somerville, Kyoungsoo Lee, HC Ferguson, and JP. Cosmic variance in the great observatories origins deep survey. *The Astrophysical*, 600(Cdm):171–174, 2004.

[14] Peter Stevenhagen. *Algebra 2*. 2010.

[15] Peter Stevenhagen. *Algebra 1*. 2016.

[16] Paul Wessel. Critical Values for the Two-sample Kolmogorov-Smirnov test (2-sided). 05(1):522, 2013.

[17] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33(2):249–274, 2005.