The Infinite Core Chase

The chase is an important tool in databases. Among the different variations of the chase, the core chase arguable has the best properties: it is the only variant that guarantees to compute a finite solution whenever one exists, and it always finds the smallest solution, called the core. In this article, we generalize the core chase to the infinite case and study its behavior when it does not terminate.

1 INFINITE CORES

Definition 1.1. A *signature* Σ is a set of relation symbols, each with an associated arity. A Σ-*structure* is a tuple (U, I) where U is a set, called the *universe*, and I is a function, called the *interpretation*, that maps each relation symbol R to a relation $I(R) \subseteq U^{ar(R)}$. We sometimes identify each relation $I(R)$ with its symbol R, and the interpretation with the set of all relations, as well as with structure itself. We write $R(a_1, \ldots, a_n)$ for the tuple $(a_1, \ldots, a_n) \in R$.

Definition 1.2. A *homomorphism* from two Σ-structures (U, I), (V, J) is a function $h : U \longrightarrow V$ such that for every relation R, if $I(a_1, \ldots, a_n) \in I$ then $R(h(a_1), \ldots, h(a_n)) \in J$. We extend a homomorphism h to tuples, relations, and structures in the natural way. In particular:

$$h(R(a_1, \ldots, a_n)) \stackrel{\text{def}}{=} (R(h(a_1), \ldots, h(a_n)))$$

$$h(R) \stackrel{\text{def}}{=} \{h(t) \mid t \in R\}$$

$$h(I) \stackrel{\text{def}}{=} \{h(R) \mid R \in I\}$$

We call the image of h the *homomorphism image*.

Definition 1.3. We call a homomorphism from a structure to itself an *endomorphism*.

Definition 1.4. Two structures I, J are *homomorphically equivalent*, written $I \equiv J$, if there are homomorphisms $h : I \longrightarrow J$ and $g : J \longrightarrow I$.

Definition 1.5. An *isomorphism* between two strucutres I, J is a function $f : I \longrightarrow J$ that has an *inverse* $f^{-1} : J \longrightarrow I$ such that $f^{-1} \circ f(I) = I$ and $f \circ f^{-1}(J) = J$. We say the structures are *isomorphic*, written $I \simeq J$.

Definition 1.6. A structure I is a *core* if every endomorphism image of I is isomorphic to I. A substructure $H \subseteq I$ is called a *core of* I if there is a homomorphism $h : I \longrightarrow H$, and H is a core.

Note that the definition of core for infinite structures is not standard, and we carefully chose our definition to ensure several desirable properties, as we shall see.

Lemma 1.7. *Two homomorphically equivalent cores are isomorphic.*

Proof. Let I, J be the cores, and $h_1 : I \longrightarrow J$, $h_2 : J \longrightarrow I$ be homomorphisms between them. Because $h_2 \circ h_1 : I \longrightarrow I$ is an endomorphism, and I is a core, $I \simeq h_2 \circ h_1(I)$. Call this isomorphism i_1, and we have $i_1 \circ h_2 \circ h_1(I) = I$. By definition $h_1(I) \subseteq J$, so $I = i_1 \circ h_2 \circ h_1(I) \subseteq i_1 \circ h_2(J)$. But $i_1 \circ h_2(J) \subseteq I$, we must have $I = i_1 \circ h_2(J)$. Similarly, we can define an inverse mapping from I to J as $J = i_2 \circ h_1$, where i_2 is an isomorphism on J defined similarly as i_1. Together, $i_1 \circ h_2$ and $i_2 \circ h_1$ define an isomorphism between I and J. ☐

Corollary 1.8. *Homomorphically equivalent structures have the same core up to isomorphism.*

Author's address:

$$\begin{tikzcd} 0 \arrow[r] & 1 \arrow[r] & 2 \arrow[r] & \cdots \end{tikzcd}$$

Fig. 1. The ray graph.

$$\begin{tikzcd} 1' & 2' & 3' & \\ 0 \arrow[r] \arrow[u] & 1 \arrow[r] \arrow[u] & 2 \arrow[r] \arrow[u] & \cdots \end{tikzcd}$$

Fig. 2. The spine graph.

Proof. Let I_1, I_2 be homomorphically equivalent structures, and let $h_1 : I_1 \to I_2$ and $h_2 : I_2 \to I_1$ be homomorphisms between them. Let H_1, H_2 be cores of I_1, I_2 respectively. Then there are homomorphisms $r_1 : I_1 \to H_1$ and $r_2 : I_2 \to H_2$, as well as $s_1 : H_1 \to I_1$ and $s_2 : H_2 \to I_2$. We can construct a pair of homomorphisms $r_2 \circ h_1 \circ s_1 : H_1 \to H_2$ and $r_1 \circ h_2 \circ s_2 : H_2 \to H_1$, showing H_1 and H_2 are homomorphically equivalent. By Lemma 1.7, they are isomorphic. $\square$

RW: Is it possible to have 2 homomorphically equivalent structures where one has a core but the other does not?

Corollary 1.9. *The core of a structure, if it exists, is unique up to isomorphism.*

1.1 Existence of the Core

Finite structures always have a core. This is however not obvious for infinite structures. Bauslaugh [3] considers several alternative definitions of core for infinite structures, and claims that there exist structures that do not have a core under any of these definitions. As we shall see in Section 2.1, our definition may not be directly comparable with Bauslaugh's definitions. We'd like to understand if all infinite structures have a core under our definition.

In the finite case, a simple method to find the core is to repeatedly apply homomorphisms until a fixpoint. However, for infinite structures, there can be an infinite sequence of homomorphisms.

Example 1.10. Consider the graph in Figure 1, taken from an example by Bauslaugh [2]. An endomorphism of this graph maps each vertex i to $i+1$, and we can repeat this mapping ad infinitum.

There can even be an infinite sequence of homomorphism images, where none of them is isomorphic to each other.

Example 1.11. Consider the graph in Figure 2. There is a sequence of homomorphisms, such that the i-th homomorphism image maps j' to j for $j \leq i$, and maps every other vertex to itself. In other words, we "fold down" one additional "spike" for every new homomorphism in the sequence. None of the homomorphism images in the sequence are isomorphic to each other. Nevertheless, the graph has a core, namely the ray graph in Figure 1.

Bodirsky [4] proved every ω-categorical structure has a substructure satisfying I and N.

Definition 1.12. A theory T (i.e., a set of sentences) is ω-categorical if any two countable models are isomorphic. A structure S is ω-categorical if its theory $T(S)$, meaning the set of all sentences satisfied by S, is ω-categorical.

RW: Can we prove a similar result for our definition of core? Or, can we prove every chase always has a model that has a core? Perhaps we can exploit the fact that the chase is specified by a *finite* set of constraints.

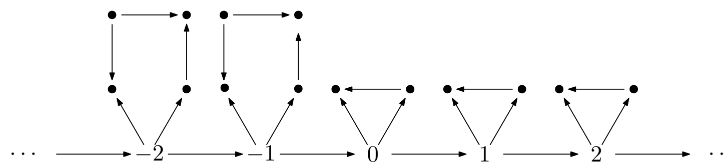

Fig. 3. The 5-3 line.

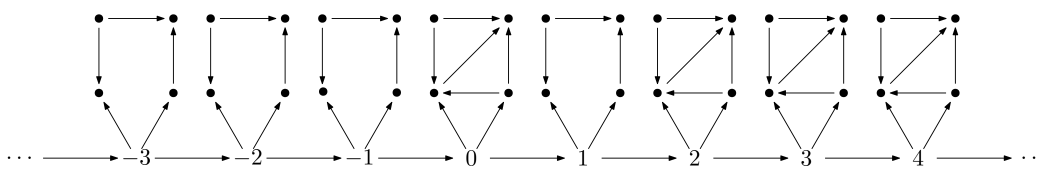

Fig. 4. The 5-line.

2 RELATED WORK

2.1 Defining Core for Infinite Structures

Bauslaugh [3] compared several different definitions of core, each one derived from the following properties:

- Property I: every endomorphism is injective.
- Property S: every endomorphism is surjective.
- Property N: every endomorphism preserves non-relations.
- Property R: there is no proper retract.

There, the notions of preserving non-relations and proper retracts are defined as follows.

Definition 2.1. A homomorphism $h : I \to J$ preserves non-relations if $R(t) \notin I \implies R(h(t)) \notin J$.

Definition 2.2. An endomorphism $h : I \to I$ is a *retraction* if $h \circ h = h$. The endomorphism image of a retraction is called a *retract*. A *proper* retract is a proper subinstance $J \subsetneq I$.

We show our definition of core is not immediately comparable with the properties above. For each property, there exist structures that are cores but do not satisfy the property, or vice versa.

Example 2.3. Consider the graph in Figure 1. Bauslaugh [2] proved every endomorphism of this graph must map vertex i to vertex $i+k$ where $k \geq 0$. Therefore the graph is a core. Yet, when $k > 0$, the endomorphism is not surjective, violating property S.

Example 2.4. Consider the graph in Figure 3. Bauslaugh [2] proved every endomorphism of the graph must map vertex i to vertex $i+k$ where $k \geq 0$, and map the polygon at i to the polygon at $i+k$. This graph is also a core, but when $k > 0$, the endomorphism is not injective, violating property I. It also does not preserve non-relations, violating property N.

Example 2.5. Consider the graph in Figure 4. Bauslaugh [2] proved an endomorphism of this graph must either be the identity; or it map vertex i to vertex $i+k$ where $k \geq 2$, where the endomorphism image is like the graph but with no edges inside the pentagon above vertex 0. In this latter case, the image is not isomorphic to the original graph, so it is not a core. Yet, the only endomorphism that is a retraction is the identity, so it satisfies property R.

RW: Can we prove our definition of core is weaker than I, S, N but stronger than R? I.e., $I \vee S \vee N \implies$ core $\implies R$? Or are they incomparable?

2.2 Extending Core Chase to the Infinite Case

There have been attempts to extend the core chase to the infinite case. Carral et al. [5] define the core in the same way as Bodirsky [4]. Baget et al. [1] do not define a core for infinite structures, but study the behavior of the core chase when it does not terminate.

RW: Can we prove the core chase converges to the infinite core?

REFERENCES

[1] Jean-François Baget, Marie-Laure Mugnier, and Sebastian Rudolph. 2023. Bounded Treewidth and the Infinite Core Chase: Complications and Workarounds toward Decidable Querying. In *Proceedings of the 42nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2023, Seattle, WA, USA, June 18-23, 2023*, Floris Geerts, Hung Q. Ngo, and Stavros Sintos (Eds.). ACM, 291–302. https://doi.org/10.1145/3584372.3588659

[2] Bruce L. Bauslaugh. 1994. Homomorphisms of infinite directed graphs. (1994).

[3] Bruce L. Bauslaugh. 1995. Core-like properties of infinite graphs and structures. *Discret. Math.* 138, 1-3 (1995), 101–111. https://doi.org/10.1016/0012-365X(94)00191-K

[4] Manuel Bodirsky. 2007. Cores of Countably Categorical Structures. *Log. Methods Comput. Sci.* 3, 1 (2007). https://doi.org/10.2168/LMCS-3(1:2)2007

[5] David Carral, Markus Krötzsch, Maximilian Marx, Ana Ozaki, and Sebastian Rudolph. 2018. Preserving Constraints with the Stable Chase. In *21st International Conference on Database Theory, ICDT 2018, March 26-29, 2018, Vienna, Austria (LIPIcs, Vol. 98)*, Benny Kimelfeld and Yael Amsterdamer (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 12:1–12:19. https://doi.org/10.4230/LIPICS.ICDT.2018.12