ON DOUBLE-MEMBERSHIP GRAPHS OF MODELS
OF ANTI-Foundation

BEA ADAM-DAY, JOHN HOWE, AND ROSARIO MENNUNI

Abstract. We answer some questions about graphs that are reducts of countable models of Anti-Foundation, obtained by considering the binary relation of double-membership $x \in y \in x$. We show that there are continuum-many such graphs, and study their connected components. We describe their complete theories and prove that each has continuum-many countable models, some of which are not reducts of models of Anti-Foundation.

This paper is concerned with the model-theoretic study of a class of graphs arising as reducts of a certain non-well-founded set theory.

Ultimately, models of a set theory are digraphs, where a directed edge between two points denotes membership. To such a model, one can associate various graphs, such as the membership graph, obtained by symmetrising the binary relation $\in$, or the double-membership graph, which has an edge between $x$ and $y$ when $x \in y$ and $y \in x$ hold simultaneously. We also consider the structure equipped with the two previous graph relations, which we call the single-double-membership graph. In [2] the first author and Peter Cameron investigated this kind of object in the non-well-founded case. We continue this line of study, and answer some questions regarding such graphs that were left open in the aforementioned work.

It is well-known that every membership graph of a countable model of ZFC is isomorphic to the Random Graph (see, e.g., [5]). The usual proof of this fact goes through for set theories much weaker than ZFC, but uses the Axiom of Foundation in a crucial way, hence the interest in (double-)membership graphs of non-well-founded set theories.

In 1917 Mirimanoff [12, 13] discussed the distinction between non-well-founded sets and their well-founded counterparts, and even presented a...
notion of isomorphism between possibly non-well-founded sets. Throughout the years they have appeared—implicitly and explicitly—in myriad places, and various formulations of axioms allowing such sets to exist have been developed and utilised. A uniform treatment of many of these axioms can be found in [1], along with historical notes.

Perhaps the most famous non-well-founded set theory is obtained from ZFC by replacing the Axiom of Foundation with the Anti-Foundation Axiom AFA, and is called ZFA (not to be confused with another ZFA, a set theory with Atoms). This axiom provides the universe with a rich class of non-well-founded sets, the structure of which reflects that of the well-founded sets: in models of ZFA there are, for example, unique $a$ and $b$ such that $a = \{b, \emptyset\}$ and $b = \{a, \{\emptyset\}\}$, and a unique $c = \{c, \emptyset, \{\emptyset\}\}$, pictured in Figure 1. By facilitating the modelling of circular behaviours, ZFA has found applications in computer science and category theory for the study of streams, communicating systems, and final coalgebras, and in philosophy, for the study of paradoxes involving circularity and natural language semantics. We refer the interested reader to [1, 3, 4].

On many accounts, models of ZFC and of ZFA are closely related, and the two set theories behave very similarly, even under forcing extensions: see for instance [7, 17]. Now, when we symmetrise the membership relation, we have two choices: we can either forget which edges were symmetric in the first place—that is, consider the membership graph—or remember this information—that is, consider the single-double-membership graph. In the first case, we find ourselves in yet another situation where the behaviour of ZFA parallels closely that of ZFC. Namely, in [2] it was proven that all membership graphs of countable models of ZFA are isomorphic to the ‘Random Loopy Graph’: the Fraïssé limit of finite graphs with self-edges. This structure is easily seen to be $\aleph_0$-categorical, ultrahomogeneous, and supersimple of SU-rank 1. If instead we take the second option, the situation changes drastically, and already double-membership graphs of models of ZFA are, in a number of senses, much more complicated. For instance, [2, Theorem 3] shows that they are not $\aleph_0$-categorical, and here we show further results in this direction.

The structure of the paper is as follows. After a brief introduction to Anti-Foundation in Section 1, and after setting up the context in Section 2,
we answer [2, Question 3] in Section 3 by characterising the connected components of double-membership graphs of models of ZFA. In the same section, we show that if we do not assume Anti-Foundation, but merely drop Foundation, then double-membership graphs can be almost arbitrary. Section 4 answers [2, Questions 1 and 2] by proving the following theorem.

**THEOREM** (Corollary 4.5). There are, up to isomorphism, continuum-many countable (single-)double-membership graphs of models of ZFA, and continuum-many countable models of each of their theories.

In Section 5 we study the common theory of double-membership graphs, which we show to be incomplete. Then, by using methods more commonly encountered in finite model theory, we characterise the completions of said theory in terms of consistent collections of consistency statements.

**THEOREM** (Theorem 5.14). The double-membership graphs of two models $M$ and $N$ of ZFA are elementarily equivalent precisely when $M$ and $N$ satisfy the same consistency statements.

We also show that all of these completions are wild in the sense of neostability theory, since each of their models interprets (with parameters) arbitrarily large finite fragments of ZFC. Our final result, below—obtained with similar techniques—answers [2, Question 5] negatively. The analogous statement for double-membership graphs holds as well.

**THEOREM** (Corollary 5.17). For every single-double-membership graph of a model of ZFA, there is a countable elementarily equivalent structure that is not the single-double-membership graph of any model of ZFA.

§1. The Anti-Foundation Axiom. There are a number of equivalent formulations of AFA. Expressed in terms of $f$-inductive functions, or of homomorphism onto transitive structures, it first appeared in [9], under the name of axiom $X_1$. It gained its current name in [1], where it was defined via decorations. The form that we shall be using is known in the literature (e.g., [4, p. 71]) as the Solution Lemma. For the equivalence with other formulations, see, e.g., [1, p. 16].

**DEFINITION** 1.1. Let $X$ be a set of ‘indeterminates’, and $A$ a set of sets. A flat system of equations is a set of equations of the form $x = S_x$, where $S_x$ is a subset of $X \cup A$ for each $x \in X$. A solution $f$ to the flat system is a function taking elements of $X$ to sets, such that after replacing each $x \in X$ with $f(x)$ inside the system, all of its equations become true.

The Anti-Foundation Axiom (AFA) is the statement that every flat system of equations has a unique solution.

**EXAMPLE** 1.2. Consider the flat system with $X = \{x, y\}$, $A = \{\emptyset, \{\emptyset\}\}$, and the following equations.
\[ x = \{ y, \emptyset \}, \]
\[ y = \{ x, \{ \emptyset \} \}. \]

The image of its unique solution \( x \mapsto a, y \mapsto b \) is pictured in Figure 1.

Note that solutions of systems need not be injective, and in fact uniqueness sometimes prevents injectivity. For instance, if \( x \mapsto a \) is the solution of the flat system consisting of the single equation \( x = \{ x \} \), then \( x \mapsto a, y \mapsto a \) solves the system with equations \( x = \{ y \} \) and \( y = \{ x \} \), whose unique solution is therefore not injective.

**Fact 1.3.** ZFC without the Axiom of Foundation proves the equiconsistency of ZFC and ZFA.

**Proof.** In one direction, from a model of ZFA one obtains one of ZFC by restricting to the well-founded sets. In the other direction, see [9, Theorem 4.2] for a class theory version, or [1, Chapter 3] for the ZFC statement.

**Remark 1.4.** There exists a weak form of AFA that only postulates the existence of solutions to flat systems, but not necessarily their uniqueness, known as axiom \( Z \) in [9] or AFA\(_1\) in [1]. Below, and in [2], uniqueness is never used; hence all the results go through for models of ZFC with Foundation replaced by AFA\(_1\). For brevity, we still state everything for ZFA.

§2. Set-up. Since Anti-Foundation allows for sets that are members of themselves, in what follows we will need to deal with graphs where there might be an edge between a point and itself. These are called loopy graphs in [2] but, for the sake of concision, we depart from common usage by adopting the following convention.

**Notation.** By graph we mean a first-order structure with a single relation that is binary and symmetric (it is not required to be irreflexive).

Since we are interested in studying (reducts of) models of ZFA, we need to assume they exist in the first place, since otherwise the answers to the questions we are studying are trivial. Therefore, in this paper we work in a set theory that is slightly stronger than usual.

**Assumption 2.1.** The ambient metatheory is ZFC + Con(ZFC).

**Definition 2.2.** Let \( L = \{ \in \} \), where \( \in \) is a binary relation symbol, and \( M \) an \( L \)-structure. Let \( S \) and \( D \) be the definable relations
\[
S(x, y) := x \in y \lor y \in x,
\]
\[
D(x, y) := x \in y \land y \in x.
\]
The single-double-membership graph, or SD-graph, \( M_0 \) of \( M \) is the reduct of \( M \) to \( L_0 := \{ S, D \} \). The double-membership graph, or D-graph, \( M_1 \) of \( M \) is the reduct of \( M \) to \( L_1 := \{ D \} \).
So, given an $L$-structure $M$, i.e., a digraph (possibly with loops) where the edge relation is $\in$, we have that $M_0 \models S(x, y)$ if and only if in $M$ there is at least one $\in$-edge between $x$ and $y$. Similarly $M_0 \models D(x, y)$ means that in $M$ we have both $\in$-edges between $x$ and $y$. The idea is that, if $M$ is a model of some set theory, then $M_0$ is a symmetrisation of $M$ that keeps track of double-membership as well as single-membership, and $M_1$ only keeps track of double-membership.

In [2], $M_0$ is called the membership graph (keeping double-edges) of $M$ and $M_1$ is called the double-edge graph of $M$. Note that, strictly speaking, SD-graphs are not graphs, according to our terminology.

For the majority of the paper we are concerned with D-graphs, since most of the results we obtain for them imply the analogous versions for SD-graphs. This situation will reverse in Theorem 5.16.

**Definition 2.3.** Let $M \models ZFA$. We say that $A \subseteq M$ is an $M$-set iff there is $a \in M$ such that $A = \{ b \in M \mid M \models b \in a \}$.

So an $M$-set $A$ is a definable subset of $M$ that is the extension of a set in the sense of $M$, namely the $a \in M$ in the definition. We will occasionally abuse notation and refer to an $M$-set $A$ when we actually mean the corresponding $a \in M$.

**§3. Connected components.** Let $M \models ZFA$. It was proven in [2, Theorem 4] that, for every finite connected graph $G$, the D-graph $M_1$ has infinitely many connected components isomorphic to $G$. It was asked in [2, Question 3] if more can be said about the infinite connected components of $M_1$. In this section we characterise them in terms of the graphs inside $M$.

Let $G$ be a graph in the sense of $M \models ZFA$, i.e., a graph whose domain and edge relation are $M$-sets, the latter as, say, a set of Kuratowski pairs. If $G$ is such a graph and $M \models \text{‘}G \text{ is connected} \text{’}$, then $G$ need not necessarily be connected. This is due to the fact that $M$ may have non-standard natural numbers; hence relations may have non-standard transitive closures. We therefore introduce the following notion.

**Definition 3.1.** Let $a \in M \models ZFA$. Let $b \in M$ be such that $M \models \text{‘}b \text{ is the transitive closure of } \{a\} \text{ under } D\text{’}$.

The **region of $a$ in $M$** is $\{ c \in M \mid M \models c \in b \}$. If $A \subseteq M$, we say that $A$ is a **region of $M$** iff it is the region of some $a \in M$.

**Remark 3.2.** For each $a \in M$, the region of $a$ in $M$ is an $M$-set.

For $a \in M$, if $A$ is the region of $a$ and $B$ is the transitive closure of $\{a\}$ under $D$ computed in the metatheory, i.e., the connected component of $a$ in $M_1$, then $B \subseteq A$. In particular, regions of $M$ are unions of connected
components of $M_1$. If $M$ contains non-standard natural numbers and the
diameter of $B$ is infinite then the inclusion $B \subseteq A$ may be strict, and $B$ may
not even be an $M$-set. From now on, the words ‘connected component’ will
only be used in the sense of the metatheory.

Most of the appeals to AFA in the rest of the paper will be applications of
the following proposition. In fact, after proving it, we will only deal directly
with flat systems twice more.

**Proposition 3.3.** Let $M_1$ be the D-graph of $M \models \text{ZFA}$, and let $G$ be a graph
in $M$. Then there is $H \subseteq M_1$ such that:

1. $(H, D^{M_1} \upharpoonright H)$ is isomorphic to $G$,
2. $H$ is a union of regions of $M$, and
3. $H$ is an $M$-set.

**Proof.** Work in $M$ until further notice. Let $G$ be a graph in $M$, say in
the language $\{ R \}$. Let $\kappa$ be its cardinality, and assume up to a suitable
isomorphism that $\text{dom } G = \kappa$. In particular, note that every element of
$\text{dom } G$ is a well-founded set. Consider the flat system

$$\{ x_i = \{ i, x_j \mid j \in \kappa, G \models R(i, j) \} \mid i \in \kappa \}.$$  

Let $s : x_i \mapsto a_i$ be a solution to the system. If $i \neq j$, then $i \in a_i \setminus a_j$, and
therefore $s$ is injective. Observe that:

(i) since $R$ is symmetric, we have $a_i \in a_j \iff G \models R(i, j)$, and
(ii) for all $b \in M$ and all $i \in \kappa$, we have $b \in a_i \in b$ if and only if there is
j $< \kappa$ such that $b = a_j$ and $G \models R(i, j)$.

Now work in the ambient metatheory. Consider the $M$-set

$$H := \{ a_i \mid M \models i \in \kappa \} = \{ b \in M \mid M \models b \in \text{Im}(s) \} \subseteq M_1.$$  

By (i) above, $(H, D^{M_1} \upharpoonright H)$ is isomorphic to $G$ and, by (ii) above, $H$ is a
union of regions of $M$. \hfill \blacksquare

We can now generalise [2, Theorem 4], answering [2, Question 3]. The
words ‘up to isomorphism’ are to be interpreted in the sense of the
metatheory, i.e., the isomorphism need not be in $M$.

**Theorem 3.4.** Let $M \models \text{ZFA}$. Up to isomorphism, the connected components
of $M_1$ are exactly the connected components (in the sense of the metatheory)
of graphs in the sense of $M$. In particular, there are infinitely many copies of
each of them.

**Proof.** Let $C$ be a connected component of a graph $G$ in $M$. By
Proposition 3.3 there is an isomorphic copy $H$ of $G$ that is a union of
regions of $M$, hence, in particular, of connected components of $M_1$. Clearly,
one of the connected components of $H$ is isomorphic to $C$. 

In the other direction, let \( a \in M_1 \) and consider its connected component. Inside \( M \), let \( G \) be the region of \( a \). Using Remark 3.2 it is easy to see that \( (G, D \upharpoonright G) \) is a graph in \( M \), and one of its connected components is isomorphic to the connected component of \( a \) in \( M_1 \).

For the last part of the conclusion take, inside \( M \), disjoint unions of copies of a given graph.

If one does not assume some form of AFA and for instance merely drops Foundation, then double-membership graphs can be essentially arbitrary, as the following proposition shows.

**Proposition 3.5.** Let \( M \models ZFC \) and let \( G \) be a graph in \( M \). There is a model \( N \) of \( ZFC \) without Foundation such that \( N_1 \) is isomorphic to the union of \( G \) with infinitely many isolated vertices, i.e., points without any edges or self-loops.

Note that the isolated vertices are necessary, as \( N \) will always contain well-founded sets.

**Proof.** Let \( G \) be a graph in \( M \), say in the language \( \{R\} \). Assume without loss of generality that \( G \) has no isolated vertices, and that \( \text{dom} \ G \) equals its cardinality \( \kappa \). For each \( i \in \kappa \) choose \( a_i \subseteq \kappa \) that has foundational rank \( \kappa \) in \( M \), e.g., let \( a_i := \kappa \setminus \{i\} \). Let \( b_j := \{a_i \mid G \models R(i, j)\} \) and note that, since no vertex of \( G \) is isolated, \( b_j \) is non-empty, and thus has rank \( \kappa + 1 \). Define \( \pi : M \to M \) to be the permutation swapping each \( a_i \) with the corresponding \( b_j \) and fixing the rest of \( M \). Let \( N \) be the structure with the same domain as \( M \), but with membership relation defined as

\[
N \models x \in y \iff M \models x \in \pi(y).
\]

By [15, Section 3]\(^1\), \( N \) is a model of ZFC without Foundation. To check that \( N_1 \) is as required, first observe that

\[
N \models a_i \in a_j \iff M \models a_i \in \pi(a_j) = b_j \iff G \models R(i, j)
\]

so \( \{a_i \mid M \models i \in \kappa\} \), equipped with the restriction of \( D^{N_1} \), is isomorphic to \( G \). To show that there are no other \( D \)-edges in \( N_1 \), assume that \( N_1 \models D(x, y) \), and consider the following three cases (which are exhaustive since \( D \) is symmetric).

(i) \( x \) and \( y \) are both fixed points of \( \pi \). This contradicts Foundation in \( M \).

(ii) \( y = a_i \) for some \( i \), so \( N \models x \in a_i \); hence \( M \models x \in \pi(a_i) = b_j \). Then \( x = a_j \) for some \( j \) by construction.

(iii) \( y = b_i \) for some \( i \). From \( N \models x \in b_i \) we get \( M \models x \in a_i \subseteq \kappa \); thus \( x \) has rank strictly less than \( \kappa \). Therefore, \( x \) is not equal to any \( a_j \) or \( b_j \); hence \( \pi(x) = x \). Again by rank considerations, it follows that \( M \models b_i \notin x = \pi(x) \), so \( N \models b_i \notin x \), a contradiction.

\(^1\)Strictly speaking, [15] works in class theory. The exact statement we use is that of [11, Chapter IV, Exercise 18].
§4. Continuum-many countable models. We now turn our attention to answering [2, Questions 1 and 2]. Namely, we compute, via a type-counting argument, the number of non-isomorphic D-graphs of countable models of ZFA and the number of countable models of their complete theories. The analogous results for SD-graphs also hold.

**Definition 4.1.** Let $n \in \omega \setminus \{0\}$. Define the $L_1$-formula

$$\varphi_n(x) := \neg D(x, x) \land \exists z_0, \ldots, z_{n-1} \left( \left( \bigwedge_{0 \leq i < j < n} z_i \neq z_j \right) \land \left( \bigwedge_{0 \leq i < n} D(z_i, x) \right) \land \left( \forall z D(z, x) \rightarrow \bigvee_{0 \leq i < n} z = z_i \right) \right).$$

For $A$ a subset of $\omega \setminus \{0\}$, define the set of $L_1$-formulas

$$\beta_A(y) := \{ \neg D(y, y) \} \cup \{ \exists x_n \varphi_n(x_n) \land D(y, x_n) \mid n \in A \} \cup \{ \neg (\exists x_n \varphi_n(x_n) \land D(y, x_n)) \mid n \in \omega \setminus (\{0\} \cup A) \}.$$

We say that $a \in M_1$ is an $n$-flower iff $M_1 \models \varphi_n(a)$. We say that $b \in M_1$ is an $A$-bouquet iff for all $\psi(y) \in \beta_A(y)$ we have $M_1 \models \psi(b)$.

So $a$ is an $n$-flower if and only if, in the D-graph, it is a point of degree $n$ without a self-loop, while $b$ is an $A$-bouquet iff it has no self-loop, it has $D$-edges to at least one $n$-flower for every $n \in A$, and it has no $D$-edges to any $n$-flower if $n \notin A$.

**Lemma 4.2.** Let $A_0$ be a finite subset of $\omega \setminus \{0\}$ and let $M \models \text{ZFA}$. Then $M_1$ contains an $A_0$-bouquet.

**Proof.** It suffices to find a certain finite graph as a connected component of $M_1$, so this follows from Proposition 3.3 (or directly from [2, Theorem 4]).

If $M$ is a structure, denote by Th($M$) its theory.

**Proposition 4.3.** Let $M \models \text{ZFA}$. Then in Th($M_1$) the $2^{\aleph_0}$ sets of formulas $\beta_A$, for $A \subseteq \omega \setminus \{0\}$, are each consistent, and pairwise contradictory. In particular, the same is true in Th($M$).

**Proof.** If $A, B$ are distinct subsets of $\omega \setminus \{0\}$ and, without loss of generality, there is an $n \in A \setminus B$, then $\beta_A$ contradicts $\beta_B$ because $\beta_A(y) \vdash \exists x_n (\varphi_n(x_n) \land D(y, x_n))$ and $\beta_B(y) \vdash \neg \exists x_n (\varphi_n(x_n) \land D(y, x_n))$.

To show that each $\beta_A$ is consistent it is enough, by compactness, to show that if $A_0$ is a finite subset of $A$ and $A_1$ is a finite subset of $\omega \setminus (\{0\} \cup A)$ then there is some $b \in M$ with a $D$-edge to an $n$-flower for every $n \in A_0$ and no $D$-edges to $n$-flowers whenever $n \in A_1$. Any $A_0$-bouquet will satisfy these requirements and, by Lemma 4.2, an $A_0$-bouquet exists inside $M_1$. 
Figure 2. The set $a = \{\{a, i\} \mid i < 5\}$ is a 5-flower. The reason for the name ‘$n$-flower’ can be seen in this figure.

For the last part, note that all the theories at hand are complete (in different languages), and whether or not an intersection of definable sets is empty does not change after adding more definable sets.

To conclude, we need the following standard fact from model theory.

**Fact 4.4.** Every partial type over $\emptyset$ of a countable theory can be realised in a countable model.

**Corollary 4.5.** Let $M$ be a model of ZFA. There are $2^{\aleph_0}$ countable models of ZFA such that their $D$-graphs (resp. $SD$-graphs) are elementarily equivalent to $M_1$ (resp. $M_0$) and pairwise non-isomorphic.

**Proof.** Consider the pairwise contradictory partial types $\beta_A$. By Fact 4.4, $\text{Th}(M)$ has $2^{\aleph_0}$ distinct countable models, as each of them can only realise countably many of the $\beta_A$. The reducts to $L_1$ (resp. $L_0$) of models realising different subsets of $\{\beta_A \mid A \subseteq \omega \setminus \{0\}\}$ are still non-isomorphic, since the $\beta_A$ are partial types in the language $L_1$.

The previous Corollary answers affirmatively [2, Questions 1 and 2].

**Remark 4.6.** For the results in this section to hold, it is not necessary that $M$ satisfies the whole of ZFA. It is enough to be able to prove Lemma 4.2 for $M$, and it is easy to see that one can provide a direct proof whenever in $M$ it is possible to define infinitely many different well-founded sets, e.g., von Neumann natural numbers, and to ensure existence of solutions to flat systems of equations. This can be done as long as $M$ satisfies Extensionality, Empty Set, Pairing, and AFA$_1$\(^2\). If we replace, in Definition 1.1, ‘$x = Sx$’ with ‘$x$ and $S_x$ have the same elements’, then we can even drop Extensionality.

\(^2\)Stated using a sensible coding of flat systems, which can be carried out using Pairing.
§5. Common theory. The main aim of this section is to study the common theory of the class of D-graphs of ZFA. We show in Corollary 5.11 that it is incomplete, and in Corollary 5.15 characterise its completions in terms of collections of consistency statements. Furthermore, we show that each of these completions is untame in the sense of neostability theory (Corollary 5.8) and has a countable model that is not a D-graph, and that the same holds for SD-graphs (Corollary 5.17), therefore solving negatively [2, Question 5].

Definition 5.1. Let $K_1$ be the class of D-graphs of models of ZFA. Let $\text{Th}(K_1)$ be its common $L_1$-theory.

Definition 5.2. Let $\varphi$ be an $L_1$-sentence. We define an $L_1$-sentence $\mu(\varphi)$ as follows. Let $x$ be a variable not appearing in $\varphi$. Let $\chi(x)$ be obtained from $\varphi$ by relativising $\exists y$ and $\forall y$ to $D(x, y)$. Let $\mu(\varphi)$ be the formula $\exists x (\neg D(x, x) \land \chi(x))$.

In other words, $\mu(\varphi)$ can be thought of as saying that there is a point whose set of neighbours is a model of $\varphi$.

Remark 5.3. Suppose $\varphi$ is a ‘standard’ sentence, i.e., one that is a formula in the sense of the metatheory, say in the finite language $L'$. Let $M \models ZFA$, and let $N$ be an $L'$-structure in $M$. Then, whether $N \models \varphi$ or not is absolute between $M$ and the metatheory. Every formula we mention is of this kind, and this fact will be used tacitly from now on.

Definition 5.4. Let $\Phi$ be the set of $L_1$-sentences that imply $\forall x, y \ (D(x, y) \rightarrow D(y, x))$.

Lemma 5.5. For every $L_1$-sentence $\varphi \in \Phi$ and every $M \models ZFA$ we have $M \models \text{Con}(\varphi) \iff M_1 \models \mu(\varphi)$.

Moreover, if this is the case, then there is $H \subseteq M_1$ such that:

1. $(H, D^{M_1} \upharpoonright H)$ satisfies $\varphi$,
2. $H$ is a union of regions of $M$, and
3. $H$ is an $M$-set.

Proof. Note that the class of graphs in $M$ is closed under the operations of removing a point or adding one and connecting it to everything. Now apply Proposition 3.3.

Define $L_{\text{NBG}} := \{E\}$, where $E$ is a binary relational symbol. We think of $L_1$ as ‘the language of graphs’ and of $L_{\text{NBG}}$ as ‘the language of digraphs’, specifically, digraphs that are models of a certain class theory (see below), hence the notation. It is well-known that every digraph is interpretable in a graph, and that such an interpretation may be chosen to be uniform, in the sense below. See, e.g., [10, Theorem 5.5.1].
**Fact 5.6.** Every $L_{NBG}$-structure $N$ is interpretable in a graph $N'$.
Moreover, for every $L_{NBG}$-sentence $\theta$ there is an $L_1$-sentence $\theta'$ such that:

1. $\theta$ is consistent if and only if $\theta'$ is, and
2. for every $L_{NBG}$-structure $N$ we have $N \models \theta \iff N' \models \theta'$.

**Corollary 5.7.** For every $L_{NBG}$-sentence $\theta$, let $\theta'$ be as in Fact 5.6. For all $M \models ZFA$,

$$M \models \text{Con}(\theta) \iff M_1 \models \mu(\theta').$$

**Proof.** Apply Lemma 5.5 to $\varphi := \theta'$.

**Corollary 5.8.** Let $M \models ZFA$. Then every model of $\text{Th}(M_1)$ interprets with parameters arbitrarily large finite fragments of ZFC. In particular $\text{Th}(M_1)$ has SOP, TP$_2$, and IP$_k$ for all $k$.

**Proof.** If $\theta$ is the conjunction of a finite fragment of ZFC, it is well-known that $ZFA \vdash \text{Con}(\theta)$. Since a model of $\theta$ is a digraph, we can apply Corollary 5.7. If $a$ witnesses the outermost existential quantifier in $\mu(\theta')$, then $\theta$ is interpretable with parameter $a$.

We now want to use Corollary 5.7 to show that the common theory $\text{Th}(K_1)$ of the class of D-graphs of models of ZFA is incomplete. Naively, this could be done by choosing $\theta$ to be a finite axiomatisation of some theory equiconsistent with ZFA, and then invoking the Second Incompleteness Theorem. For instance, one could choose von Neumann–Bernays–Gödel class theory NBG, axiomatised in the language $L_{NBG}$, as this is known to be equiconsistent with ZFC (see [8]), hence with ZFA. The problem with this argument is that, in order for it to work, we need a further set-theoretical assumption in our metatheory, namely $\text{Con}(ZFC + \text{Con}(ZFC))$. This can be avoided by using another sentence whose consistency is independent of ZFA, provably in $ZFC + \text{Con}(ZFC)$ alone. We would like to thank Michael Rathjen for pointing out to us the existence of such a sentence.

Let NBG$^-$ denote NBG without the axiom of Infinity. We will use special cases of a classical theorem of Rosser and of a related result. For proofs of these, together with their more general statements, we refer the reader to [16, Chapter 7, Application 2.1 and Corollary 2.6].

**Fact 5.9** (Rosser’s Theorem). There is a $\Pi_1^0$ arithmetical statement $\psi$ that is independent of ZFA.

**Fact 5.10.** Let $\psi$ be a $\Pi_1^0$ arithmetical statement. There is another arithmetical statement $\tilde{\psi}$ such that $ZFA \vdash \psi \leftrightarrow \text{Con}(\text{NBG}^- + \tilde{\psi})$.

**Corollary 5.11.** $\text{Th}(K_1)$ is not complete.

---

3 The reader may have encountered an axiomatisation using two sorts; this can be avoided by declaring sets to be those classes that are elements of some other class.
Proof. Let $\psi$ be given by Rosser’s Theorem, and let $\tilde{\psi}$ be given by Fact 5.10 applied to $\psi$. Apply Corollary 5.7 to $\theta := \text{NBG}^- + \tilde{\psi}$. ⊢

It is therefore natural to study the completions of $\text{Th}(K_1)$, and it follows easily from $K_1$ being pseudoelementary that all of these are the theory of some actual D-graph $M_1$. We provide a proof for completeness.

**Proposition 5.12.** Let $T$ be an $L$-theory, and let $K$ be the class of its models. Let $L_1 \subseteq L$, and for $M \in K$ denote $M_1 := M \upharpoonright L_1$. Let $K_1 := \{ M_1 | M \in K \}$ and $N \models \text{Th}(K_1)$. Then there is $M \in K$ such that $M_1 \equiv N$.

Proof. We are asking whether there is any $M \models T \cup \text{Th}(N)$, so it is enough to show that the latter theory is consistent. If not, there is an $L_1$-formula $\varphi \in \text{Th}(N)$ such that $T \vdash \neg \varphi$. In particular, since $\neg \varphi \in L_1$, we have that $\text{Th}(K_1) \vdash \neg \varphi$, and this contradicts that $N \models \text{Th}(K_1)$. ⊢

In order to characterise the completions of $\text{Th}(K_1)$, we will use techniques from finite model theory, namely Ehrenfeucht–Fraïssé games and $k$-equivalence. For background on these concepts, see [6].

**Lemma 5.13.** Let $G = G_0 \sqcup G_1$ be a graph with no edges between $G_0$ and $G_1$, and let $H = H_0 \sqcup H_1$ be a graph with no edges between $H_0$ and $H_1$. If $(G_0, a_1, \ldots, a_{m-1}) \equiv_k (H_0, b_1, \ldots, b_{m-1})$ and $(G_1, a_m) \equiv_k (H_1, b_m)$, then $(G, a_1, \ldots, a_m) \equiv_k (H, b_1, \ldots, b_m)$.

Proof. This is standard, see, e.g., [6, Proposition 2.3.10]. ⊢

**Theorem 5.14.** Let $M$ and $N$ be models of ZFA. The following are equivalent.

1. $M_1 \equiv N_1$.
2. $M_1$ and $N_1$ satisfy the same sentences of the form $\mu(\varphi)$, as $\varphi$ ranges in $\Phi$.
3. $M$ and $N$ satisfy the same consistency statements.

Proof. For statements about graphs, the equivalence of 2 and 3 follows from Lemma 5.5. For statements in other languages, it is enough to interpret them in graphs using [10, Theorem 5.5.1].

For the equivalence of 1 and 2, we show that for every $n \in \omega$ the Ehrenfeucht–Fraïssé game between $M_1$ and $N_1$ of length $n$ is won by the Duplicator, by describing a winning strategy. The idea behind the strategy is the following. Recall that, for every finite relational language and every $k$, there is only a finite number of $\equiv_k$-classes, each characterised by a single sentence (see, e.g., [6, Corollary 2.2.9]). After the Spoiler plays a point $a$, the Duplicator replicates the $\equiv_k$-class of the region of $a$ using Lemma 5.5.

Fix the length $n$ of the game and denote by $a_1, \ldots, a_m \in M_1$ and $b_1, \ldots, b_m \in N_1$ the points chosen at the end of turn $m$. The Duplicator defines, by simultaneous induction on $m$, sets $G_0^m \subseteq M_1$ and $H_0^m \subseteq N_1$, and makes sure that they satisfy the following conditions.
Before the game starts (‘after turn 0’) we set $G^0_0 = H^0_0 = \emptyset$ and all conditions trivially hold. Assume inductively that they hold after turn $m$ – 1. We deal with the case where the Spoiler plays $a_m \in M_1$: the case where the Spoiler plays $b_m \in N_1$ is symmetrical.

Let $G^m_1$ be the region of $a_m$ in $M$. If $G^m_1 \subseteq G^{m-1}_0$ then, by inductive hypothesis condition (C4) held after turn $m$ – 1, the Duplicator can find $b_m \in H^{m-1}_0$ such that $(G^{m-1}_0, a_0, \ldots, a_m) \equiv_{n-m} (H^{m-1}_0, b_0, \ldots, b_m)$. It is then clear that all conditions hold after setting $G^m_0 = G^{m-1}_0$ and $H^m_0 = H^{m-1}_0$.

Otherwise, by (C2), we have $G^m_1 \cap G^{m-1}_0 = \emptyset$. Let $\varphi$ characterise the $\equiv_{n-m+1}$-class of $G^m_1$. Note that, if $n - m + 1 \geq 2$, then $\varphi \in \Phi$ automatically. Otherwise, replace $\varphi$ with $\varphi \land \forall x \forall y (D(x, y) \rightarrow D(y, x))$. By Remark 3.2, $G^m_1$ is an $M$-set, hence $M \models Con(\varphi)$. By Lemma 5.5 and assumption, there is a union $H^m_1$ of regions of $N$ which is an $N$-set and such that $G^m_1 \equiv_{n-m+1} H^m_1$. By inductive hypothesis, $H^{m-1}_0$ is also an $N$-set by (C3). Therefore, up to writing a suitable flat system in $N$, we may replace $H^m_1$ with an isomorphic copy that is still a union of regions and an $N$-set, but with $H^m_1 \cap H^{m-1}_0 = \emptyset$.

Let $b_m \in H^m_1$ be the choice given by a winning strategy for the Duplicator in the game of length $n - m + 1$ between $G^m_1$ and $H^m_1$ after the Spoiler plays $a_m \in G^m_1$ as its first move. Set $G^m_0 = G^{m-1}_0 \cup G^m_1$ and $H^m_0 = H^{m-1}_0 \cup H^m_1$. Note that $G^{m-1}_0, G^m_1, H^{m-1}_0, H^m_1$ are all unions of regions and $M$-sets or $N$-sets; hence (C2) and (C3) hold (and (C1) is clear). Moreover both unions are disjoint, so the hypotheses of Lemma 5.13 are satisfied and $(G^m_0, a_1, \ldots, a_m) \equiv_{n-m} (H^m_0, b_1, \ldots, b_m)$, i.e., (C4) holds.

To show that this strategy is winning, note that the outcome of the game only depends on the induced structures on $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ at the end of the final turn. These do not depend on what is outside $G^m_0$ and $H^m_0$ since they are unions of regions, hence unions of connected components. As (C4) holds at the end of turn $n$, the structures induced on $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are isomorphic.

**Corollary 5.15.** Let $N \models Th(K_1)$. Then $Th(N)$ is axiomatised by

$$Th(K_1) \cup \{\mu(\varphi) \mid \varphi \in \Phi, N \models \mu(\varphi)\} \cup \{\neg \mu(\varphi) \mid \varphi \in \Phi, N \models \neg \mu(\varphi)\}.$$ 

**Proof.** Let $N'$ satisfy the axiomatisation above. Since $N$ and $N'$ are models of $Th(K_1)$ we may, by Proposition 5.12, replace them with $D$-graphs $M_1 \equiv N$ and $M'_1 \equiv N'$ of models of ZFA. By Theorem 5.14 $M_1 \equiv M'_1$. 

$\square$
By the previous corollary, combined with Lemma 5.5, theories of double-membership graphs correspond bijectively to consistent (with ZFA, equivalently with ZFC) collections of consistency statements.

The reader familiar with finite model theory may have noticed similarities between the proof of Theorem 5.14 and certain proofs of the theorems of Hanf and Gaifman (see [6, Theorems 2.4.1 and 2.5.1]). In fact one could deduce a statement similar to Theorem 5.14 directly from Gaifman’s Theorem. This would characterise the completions of \( \text{Th}(K_1) \) in terms of \textit{local formulas}, of which the \( \mu(\varphi) \) form a subclass, yielding a less specific result than Corollary 5.15. Moreover, we believe that the correspondence with collections of consistency statements provides a conceptually clearer picture.

Similar ideas can be used to study [2, Question 5], which asks whether a countable structure elementarily equivalent to the SD-graph \( M_0 \) of some \( M \models \text{ZFA} \) must itself be the SD-graph of some model of ZFA. We provide a negative solution in Corollary 5.17. Again, Gaifman’s Theorem could be used directly to deduce its second part.

\textbf{Theorem 5.16.} Let \( M \models \text{ZFA} \). There is a countable \( N \equiv M_0 \) such that \( N \upharpoonright L_1 \) has no connected component of infinite diameter.

Before the proof, we show how this solves [2, Question 5].

\textbf{Corollary 5.17.} For every \( M \models \text{ZFA} \) there are a countable \( N \equiv M_0 \) that is not the SD-graph of any model of ZFA and a countable \( N' \equiv M_1 \) that is not the D-graph of any model of ZFA.

\textbf{Proof.} Let \( N \) be given by Theorem 5.16 and \( N' := N \upharpoonright L_1 \). Now observe that, as follows easily from Proposition 3.3, any reduct to \( L_1 \) of a model of ZFA has a connected component of infinite diameter.

Note that this proves slightly more: a negative solution to the question would only have required to find a single pair \( (M_0, N) \) satisfying the conclusion of the corollary.

\textbf{Proof of Theorem 5.16.} Up to passing to a countable elementary substructure, we may assume that \( M \) itself is countable. Let \( N \) be obtained from \( M_0 \) by removing all points whose connected component in \( M_1 \) has infinite diameter. We show that \( M_0 \equiv N \) by exhibiting, for every \( n \), a sequence \( (I_j)_{j \leq n} \) of non-empty sets of partial isomorphisms between \( M_0 \) and \( N \) with the back-and-forth property (see [6, Definition 2.3.1 and Corollary 2.3.4]). The idea is to adapt the proof of [14, Lemma 2.2.7] (essentially Hanf’s Theorem) by considering the Gaifman balls with respect to \( L_1 \), while requiring the partial isomorphisms to preserve the richer language \( L_0 \).

On an \( L_0 \)-structure \( A \), consider the distance \( d : A \to \omega \cup \{ \infty \} \) given by the graph distance in the reduct \( A \upharpoonright L_1 \) (where \( d(a,b) = \infty \) iff \( a,b \) lie in
distinct connected components). If \( a_1, \ldots, a_k \in A \) and \( r \in \omega \), denote by \( \text{dom}(B(r, a_1, \ldots, a_k)) \) the union of the balls of radius \( r \) (with respect to \( d \)) centred on \( a_1, \ldots, a_k \). Equip \( \text{dom}(B(r, a_1, \ldots, a_k)) \) with the \( L_0 \)-structure induced by \( A \), then expand to an \( L_0 \cup \{ c_1, \ldots, c_k \} \)-structure \( B(r, a_1, \ldots, a_k) \) by interpreting each constant symbol \( c_i \) with the corresponding \( a_i \). We stress that, even though \( B(r, a_1, \ldots, a_k) \) carries an \( L_0 \cup \{ c_1, \ldots, c_k \} \)-structure, and we consider isomorphisms with respect to this structure, the balls giving its domain are defined with respect to the distance induced by \( L_1 \) alone.

Set \( r_j := (3^j - 1)/2 \) and fix \( n \). Define \( I_0 := \{ \emptyset \} \), where \( \emptyset \) is thought of as the empty partial map \( M_0 \to N \). For \( j < n \), let \( I_j \) be the following set of partial maps \( M_0 \to N \):

\[
I_j := \{ a_1, \ldots, a_k \mapsto b_1, \ldots, b_k \mid k \leq n - j, B(r_j, a_1, \ldots, a_k) \cong B(r_j, b_1, \ldots, b_k) \}.
\]

We have to show that for every map \( a_1, \ldots, a_k \mapsto b_1, \ldots, b_k \) in \( I_{j+1} \) and every \( a \in M_0 \) [resp. every \( b \in N \)] there is \( b \in N \) [resp. \( a \in M_0 \)] such that \( a_1, \ldots, a_k, a \mapsto b_1, \ldots, b_k, b \) is in \( I_j \).

Denote by \( i \) an isomorphism \( B(r_{j+1}, a_1, \ldots, a_k) \to B(r_{j+1}, b_1, \ldots, b_k) \) and let \( a \in M_0 \). If \( a \) is chosen in \( B(2 \cdot r_{j+1}, a_1, \ldots, a_k) \), then by the triangle inequality and the fact that \( 2 \cdot r_{j+1} + r_{j+1} = r_{j+1} \) we have \( B(r_j, a) \subseteq B(r_{j+1}, a_1, \ldots, a_k) \), and we can just set \( b := i(a) \).

Otherwise, again by the triangle inequality, \( B(r_j, a) \) and \( B(r_j, a_1, \ldots, a_k) \) are disjoint and there is no \( D \)-edge between them. Note, moreover, that they are \( M \)-sets. This allows us to write a suitable flat system, which will yield the desired \( b \).

Working inside \( M \), for every \( d \in B(r_j, a) \) choose a well-founded set \( h_d \) such that for all \( d_0, d_1 \in B(r_j, a) \) we have:

\[
\begin{align*}
\text{(H1)} & \quad h_{d_0} \notin h_{d_1}, \\
\text{(H2)} & \quad \text{if } d_0 \neq d_1 \text{ then } h_{d_0} \neq h_{d_1}, \\
\text{(H3)} & \quad h_d \notin B(r_j, b_1, \ldots, b_k), \\
\text{(H4)} & \quad h_d \notin \bigcup B(r_j, b_1, \ldots, b_k), \quad \text{and} \\
\text{(H5)} & \quad h_d \notin \bigcup B(r_j, b_1, \ldots, b_k). 
\end{align*}
\]

Let \( \{ x_d \mid d \in B(r_j, a) \} \) be a set of indeterminates. Define

\[
\begin{align*}
P_d & := \{ x_e \mid e \in B(r_j, a), M \models e \in d \}, \\
Q_d & := \{ i(f) \mid f \in B(r_j, a_1, \ldots, a_k), M \models S(d, f) \},
\end{align*}
\]

and consider the flat system

\[
\{ x_d = \{ h_d \} \cup P_d \cup Q_d \mid d \in B(r_j, a) \}.
\]

Intuitively, the terms \( P_d \) ensure that the image of a solution is an isomorphic copy of \( B(r_j, a) \), while the terms \( Q_d \) create the appropriate \( S \)-edges between the image and \( B(r_j, b_1, \ldots, b_k) \) (note that we do not need any \( D \)-edges
because there are none between \( B(r_j, a) \) and \( B(r_j, a_1, \ldots, a_k) \). The \( \{h_d\} \) are needed for bookkeeping reasons, in order to avoid pathologies. We now spell out the details; keep in mind that each \( P_d \) consists of indeterminates, and each \( Q_d \) is a subset of \( B(r_j, b_1, \ldots, b_k) \).

Let \( s \) be a solution of \((*)\), guaranteed to exist by AFA. By \((H1)\) and the fact that each member of \( \text{Im}(s) \) contains some \( h_d \), we have \( \{h_d \mid d \in B(r_j, a)\} \cap \text{Im}(s) = \emptyset \). Using this together with \((H2)\) and \((H3)\) we have \( h_d \in s(x_e) \iff d = e \); hence \( s \) is injective.

Let \( s' := d \mapsto s(x_d) \) and \( b := s'(a) \). By \((H4)\) we have that \( \text{Im}(s) \) does not intersect \( B(r_j, b_1, \ldots, b_k) \), and we already showed that it does not meet \( \{h_d \mid d \in B(r_j, a)\} \). By looking at \((*)\) and at the definition of the terms \( P_d \), we have that \( \text{Im}(s) = B(r_j, b) \) and that \( s' \) is an isomorphism \( B(r_j, a) \to B(r_j, b) \).

Note that the only \( D \)-edges involving points of \( \text{Im}(s) \) can come from the terms \( P_d \): the \( h_d \) are well-founded, and there are no \( g \in \text{Im}(s) \) and \( \ell \in B(r_j, b_1, \ldots, b_k) \) such that \( g \in \ell \), since \( g \) contains some \( h_j \) but this cannot be the case for any element of \( \ell \) because of \((H5)\). Hence \( \text{Im}(s) \) is a connected component of \( M_1 \) and it has diameter not exceeding \( 2 \cdot r_j \), so is included in \( N \).

Set \( \iota' := s' \cup (t \upharpoonright B(r_j, a_1, \ldots, a_k)) \). This map is injective because it is the union of two injective maps whose images \( B(r_j, b) \) and \( B(r_j, b_1, \ldots, b_k) \) are, as shown above, disjoint. Moreover, there are no \( D \)-edges between \( B(r_j, b) \) and \( B(r_j, b_1, \ldots, b_k) \), since the former is a connected component of \( M_1 \). By inspecting the terms \( Q_d \), we conclude that \( \iota' \) is an isomorphism \( B(r_j, a_1, \ldots, a_k, a) \to B(r_j, b_1, \ldots, b_k, b) \), and this settles the ‘forth’ case.

The proof of the ‘back’ case, where we are given \( b \in N \) and need to find \( a \in M_0 \), is analogous (and shorter, as we do not need to ensure that the new points are in \( N \)); we can consider statements such as \( e \in d \) when \( e, d \in N \) since the domain of the \( L_0 \)-structure \( N \) is a subset of \( M \).

PROBLEMS. We leave the reader with some open problems.

1. Axiomatise the theory of \( D \)-graphs of models of ZFA.
2. Axiomatise the theory of \( SD \)-graphs of models of ZFA.
3. Characterise the completions of the theory of \( SD \)-graphs of models of ZFA.

Acknowledgements. We are grateful to Michael Rathjen for pointing out to us Fact 5.10, and to Dugald Macpherson and Vincenzo Mantova for their guidance and feedback. The first author is supported by a Leeds Doctoral Scholarship. The second and third authors were supported by Leeds Anniversary Research Scholarships. The third author is supported by the project PRIN 2017: “Mathematical Logic: models, sets, computability” Prot. 2017NWTM8RPRIN.
REFERENCES

[1] P. Aczel, *Non-Well-Founded Sets*. CSLI Lecture Notes, vol. 14, CSLI Publications, Stanford, 1988.

[2] B. Adam-Day and P. J. Cameron, *Undirecting membership in models of anti-foundation*. *Aequationes Mathematicae*, vol. 95 (2021), pp. 393–400.

[3] J. Barwise and J. Etchemendy, *The Liar: An Essay on Truth and Circularity*. Oxford University Press, New York and Oxford, 1987.

[4] J. Barwise and L. S. Moss, *Vicious Circles*. CSLI Lecture Notes, vol. 60, CSLI Publications, Stanford, 1996.

[5] P. Cameron, *The random graph*. *The Mathematics of Paul Erdős*, vol. 2 (R. L. Graham, J. Nešetřil, and S. Butler, editors). Springer, Berlin and New York, 2013, pp. 353–378.

[6] H.-D. Ebbinghaus and J. Flum, *Finite Model Theory*. Springer Monographs in Mathematics. Springer, 1995.

[7] O. Esser, *Forcing with the anti-foundation axiom*. *Mathematical Logic Quarterly*, vol. 58 (2012), pp. 55–62.

[8] U. Felgner, *Comparison of the axioms of local and universal choice*. *Fundamenta Mathematicae*, vol. 71 (1971), pp. 43–62.

[9] M. Forti and F. Honsell, *Set theory with free construction principles*. *Annali della Scuola Normale Superiore di Pisa—Classe di Scienze*, vol. 10 (1983), pp. 493–522.

[10] W. Hodges, *Model Theory*, Encyclopedia of Mathematics and its Applications, vol. 42, Cambridge University Press, Cambridge and New York, 1993.

[11] K. Kunen, *Set Theory*. Studies in Logic and the Foundations of Mathematics, vol. 102, North-Holland, Amsterdam, New York, and Oxford, 1980.

[12] D. Mirimanoff, *Les antinomies de Russell et de Burali–Forti et le probleme fondamental de la theorie des ensembles*. *L'enseignement mathematique*, vol. 19 (1917), pp. 37–52.

[13] ———. *Remarques Sur la theorie des ensembles*. *L'enseignement mathematique*, vol. 19 (1917), pp. 209–217.

[14] M. Otto, *Finite model theory*. 2006. Available at https://www.karlin.mff.cuni.cz/~krajicek/otto.pdf.

[15] L. Rieger, *A contribution to Gödel’s axiomatic set theory. I*. *Czechoslovak Mathematical Journal*, vol. 7 (1957), pp. 323–357.

[16] C. Smoryński, *Self-Reference and Modal Logic*. Universitext. Springer, New York, 1985.

[17] M. Viale, *The cumulative hierarchy and the constructible universe of ZFA*. *Mathematical Logic Quarterly*, vol. 50 (2004), pp. 99–103.

SCHOOL OF MATHEMATICS
UNIVERSITY OF LEEDS
LEEDS, UK
E-mail: beadamday@gmail.com
E-mail: johnhowe82@hotmail.co.uk

DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI PISA
PISA, ITALY
E-mail: R.Mennuni@posteo.net