INFINITE LIMITS OF COPYING MODELS OF THE WEB GRAPH

ANTHONY BONATO AND JEANNETTE JANSSEN

Abstract. Several stochastic models were proposed recently to model the dynamic evolution of the web graph. We study the infinite limits of the stochastic processes proposed to model the web graph when time goes to infinity. We prove that deterministic variations of the so-called copying model can lead to several non-isomorphic limits. Some models converge to the infinite random graph R, while the convergence of other models is sensitive to initial conditions or minor changes in the rules of the model. We explain how limits of the copying model of the web graph share several properties with R that seem to reflect known properties of the web graph.

1. Introduction

The web may be viewed as a directed graph with nodes the static HTML web pages, and directed edges representing the links between web pages. This graph is commonly referred to as *the web graph*; it is an example of a massive network, with several billion nodes. Several interesting properties were observed in the web graph: in particular, the in- and out-degrees seem to satisfy a power law degree distribution, the web graph is *small world*, which means that it has high clustering and low diameter, and it is locally dense while globally sparse. (See [20] for a survey of properties of the web graph.) Another interesting property of the web graph is that it is *evolutionary*: nodes appear and disappear with time. Throughout this paper, we will consider the simple, undirected version of the web graph. (The reason for this is that the structural results we present are best described in graphs where edges have no orientation; we consider the directed case as the next step in our study.)

1991 *Mathematics Subject Classification.* 05C80, 68R10, 94C15.

Key words and phrases. web graph, evolving copying model, preferential attachment model, the infinite random graph, adjacency property, inexhaustible graph, bipartite cores, Hamilton paths.

The authors gratefully acknowledges support from NSERC research grants, and from a MITACS grant.

Owing to its massive and dynamic nature, several authors have suggested statistical models which capture certain properties of the web graph. These models are loosely based on classical random graphs, first introduced by Erdős and Rényi. If n is a positive integer, and $0 < p < 1$ is a fixed real number, then a random graph $G(n, p)$ has n nodes, and there is an edge between two nodes with probability p. The graphs $G(n, p)$ have several drawbacks as models of the web graph. For example, the degree distribution of random graphs is binomial, rather than satisfying a power law; further, the number of nodes is static. These drawbacks maybe overcome by making the model dynamic, and by assigning different probabilities to various nodes. Two models that take these approaches are the *preferential attachment model* of Barabási and Albert [3] and the *evolving copying model* of Kumar et at [19]. In the preferential attachment model, we start with a small base graph. At each time step, we create a new node, say u, and draw its edges according to a predetermined distribution. In particular, node u is joined to an existing node v with probability proportional to $\deg(v)$. In the evolving copying model, we start with a small base graph. At each time step, we create a new node, u. Choose an existing node v uniformly at random (u.a.r.). For each edge vw, with probability $1 - p$, add the edge uw. Hence, the neighbourhood of the new node u will be a subset of the neighbourhood of the existing node v. A slight variation is the *evolving copying model with error*, where with probability p, an edge is added between u and an existing node chosen u.a.r.

The first analysis of the long-term behaviour of these models has been made, for example, by Aiello, Chung, Lu [1, 2], Cooper, Frieze [10], and by Kumar *et al.* [19]. Power law degree distributions were proven to exist in both the preferential attachment [3] and evolving copying models [19]. Many bipartite cliques were shown to exist in evolving copying models [19], mirroring the abundance of so-called "cyber-communities" measured by bipartite cliques in the web graph (as reported in [18]).

Our motivating question is: *what are the resulting graphs like if we allow these stochastic processes to continue indefinitely*? We attempt to answer this question in the case of the copying model of the web graph. The resulting graphs are infinite, and are *limits* (that is, unions of chains) of finite web graph models. On the surface, the study of infinite graphs may appear to have no connection with the study of a finite experimental graph such as the web graph. However, limits of web graph models have certain fractal and other properties which correlate with known data on the web graph obtained by various web crawls. (See Theorems 7 and 9.)

If we consider limits of the $G(n, p)$ graphs, then the resulting graph will almost surely be isomorphic to the *infinite random graph*, written R. The graph R is the unique (up to isomorphism) countable graph satisfying the following *existentially closed* or *e.c.* adjacency property.

e.c. property: A graph G is e.c. if for each pair of finite disjoint subsets X and Y of nodes of G, there exists a node $z_{X,Y} \in V(G) \setminus (X \cup Y)$ that is joined to each node of X and to no node of Y.

For more on R, the reader is directed to the excellent survey [9]. The graph R may be viewed as the limit of an evolutionary process. For this, let R_0 be a single node; assume that R_n is defined and contains R_0. Enumerate all of the finite subsets of nodes of R_n, and extend each of these subsets, in all possible ways, by new nodes not in R_n. The resulting graph we call R_{n+1}, and the union of the chain $(R_n : n \in \omega)$ is an e.c. graph that is isomorphic to R. The preceding construction of R serves as a template for what follows, where we will consider infinite graphs grown by certain evolutionary processes. Our results show that graphs grown in this way have many properties in common with R, although they are usually not isomorphic to R. See Sections 3 and 4.

2. Adjacency properties and limits

All the graphs we consider are undirected, simple, and have a countable number of nodes. We use the notation ω for the set of natural numbers considered as an ordinal, and $\aleph_0$ is the cardinality of ω. The cardinality of the real numbers is written $2^{\aleph_0}$. If $S \subseteq V(G)$, then $G \restriction S$ is the subgraph induced by S. If G is an induced subgraph of H, then we write $G \leq H$. The graph $G \uplus H$ is the disjoint union of G and H. If y is a node of G, then $N(y) = \{z : yz \in E(G)\}$ is the *neighbour set* of y in G. The *closed neighbour set* of y, written $N[y]$, is the set $N(y) \cup \{y\}$. If x is a node of G, then the graph $G - x$ is the graph $G \restriction (V(G) \setminus \{x\})$. If $S \subseteq V(G)$, then $G - S$ is defined similarly. A node is *isolated* if it has no neighbours, and it is *universal* if it joined to all nodes except itself.

To study the limits of web graph models, we consider graphs satisfying various deterministic adjacency properties that are more general than the e.c. property described in the Introduction. We note that determinism has been used by other authors in the study of the web graph; for instance, see [4] for a study of deterministic scale-free networks. Let X and Y be disjoint finite sets of nodes in a graph G. We say that the node $z_{X,Y} \in V(G) \setminus (X \cup Y)$ is *correctly joined to* X and Y, if $z_{X,Y}$ is joined to each node of X and no node of Y.

Property (A): A graph G has property (A) if for each node y of G, for each finite $X \subseteq N[y]$, and each finite $Y \subseteq V(G) \setminus X$, there exists a node $z_{X,Y} \neq y$ which is correctly joined to X and Y.

Property (B): Property (B) is defined similarly to Property (A), except that $N[y]$ is replaced by $N(y)$.

For a fixed $n \in \omega$, properties (A,n) and (B,n) are defined analogously to (A) and (B) respectively, but the node z may be joined to at most n other nodes. More precisely:

Property (A,n): A graph G has property (A,n) for some $n \in \omega$ if for each node y of G, for each finite $X \subseteq N[y]$, for each finite $Y \subseteq V(G) \setminus X$, and for each set $U \subseteq V(G) \setminus (N[y] \cup Y)$ with cardinality at most n, there is a node $z_{X,Y,U} \neq y$ correctly joined to $X \cup U$ and Y.

Property (B,n):
Again, Property (B,n) is defined similarly to Property (A,n), except that $N[y]$ is replaced by $N(y)$.

Note that property (A) is just (A,0), and (B) is just (B,0). We sometimes say that a graph with property ($\mathcal{P}$), where $\mathcal{P}$ is one of A or B, is a *($\mathcal{P}$) graph*.

The adjacency properties (A) and (B) are inspired by the evolving copying model of the web graph, while properties (A,n) and (B,n) are inspired by the evolving copying model with error. The idea (that will be made precise in the sequel) is that as time goes to infinity, any extension that is made with positive probability is almost surely true in the limit.

Is there anything that can be said about the structure of graphs with these adjacency properties? How do these graphs compare and contrast with R, and with the actual web graph? In this section and the next, we will attempt to answer these questions. A first observation is that we have the following chain of logical implications (for all integers $n \geq 1$),

$$\text{e.c.} \Rightarrow (\text{A},n) \Rightarrow (\text{B},n) \Rightarrow (\text{A}) \Rightarrow (\text{B}).$$

Our first theorem gives insight into the structure of graphs with (A). A graph is $\aleph_0$-*universal* if it embeds all countable graphs as induced subgraphs. For example, it well-known that R is $\aleph_0$-universal; see [9], for example.

Theorem 1. *Let G satisfy (A). Then for all $y \in V(G)$, $G \restriction N(y) \cong R$. In particular, G is $\aleph_0$-universal.*

Proof. Fix $y \in V(G)$. By remarks in the Introduction, it is enough to show that $N = G \restriction N(y)$ is e.c. For this, fix X and Y, disjoint subsets of $V(N)$. By (A) there is a node $z_{X,Y}$ of G that is joined to each node of $\{y\} \cup X$, but is not joined nor equal to any of the nodes in Y. Then $z_{X,Y}$ is a node of N, and therefore, N satisfies the e.c. property. □

In the next theorem, we see that if $n \geq 1$, then the adjacency properties (A,n) and (B,n) are in fact *equivalent* to the e.c. property. We find this surprising, since then adding a single extra "random link" gives a deterministic conclusion. As we will see in Theorem 5, however, there are uncountably many non-isomorphic countable graphs with (A) or (B). Thus, Theorems 2 and 5 seem to contrast the evolving copying and evolving copying with error models.

Theorem 2. *Fix $n > 0$ an integer. If G has (B,n), then G is isomorphic to R.*

Proof. To show that G is isomorphic to R, we need only show that G is e.c.: for all finite disjoint subsets C and D of nodes of G, there is a node z of $V(G) \setminus (C \cup D)$ that is joined to each node of C and to no node of D.

Case 1. $|C| \leq n$.

Choose any node $y \notin C \cup D$. Let $X = C \cap N(y)$, $Y = D$, and $U = C \setminus N(y)$. By property (B,n), there exists a node $z_{X,U,Y}$ correctly joined to $X \cup U = C$ and $Y = D$.

Case 2. $|C| > n$.

We can then write C as $W_1 \cup \ldots \cup W_r$, where the W_i are pairwise disjoint and have cardinality at most n. Now, as in Case 1, choose a node y not in $C \cup D$. Let $X = N(y) \cap W_1$, $Y = D$, and let $U = W_1 \setminus N(y)$. By the (B,n), there is a node $x_1 \notin X \cup Y$ such that $W_1 \subseteq N(x_1)$. By the (B,n) property with $X = W_1$, Y empty, and $U = W_2$, there is a node x_2 not in $X \cup Y$ that is joined to all of $W_1 \cup W_2$. Proceeding inductively, we can find a node x_r that is joined to all of the nodes in $C \cup D$. As then $C \cup D \subseteq N(x_r)$, by a final application of (B,n), there is a node z correctly joined to C and D. □

Because of Theorem 2, we will restrict our attention to graphs satisfying properties (A)=(A,0) and (B)=(B,0) for the rest of the paper. If $(G_n : n \in \omega)$ is a sequence of graphs with $G_n \leq G_{n+1}$, then define

$$\lim_{n\to\infty} G_n = \bigcup_{n\in\omega} G_n;$$

we call $\lim_{n\to\infty} G_n$ the *limit* of the sequence $(G_n : n \in \omega)$. We say that $(G_n : n \in \omega)$ is a *chain* of graphs.

A graph G has an (A)-*constructing sequence* if there is a chain of graphs $(G_n : n \in \omega)$ such that $lim_{n\rightarrow\infty} G_n = G$, which has the following properties. (A (B)-*constructing sequence* is defined analogously, using $N(y)$ rather than $N[y]$.)

(1) G_0 is a finite graph.
(2) For each integer $n > 1$, G_{n+1} is obtained from G_n by one application of process *(P1)* followed by a finite number (possibly zero) applications of process *(P2)*, where (P1) and (P2) are defined as follows:
  (P1) For each node y of G_n, and for each finite $X \subseteq N[y]$, a new node $z_X \notin V(G_n)$ is added whose neighbours in $V(G_n)$ are exactly the nodes of X. We say that z_X *extends* X. We say that *all subsets of closed neighbour sets in* $V(G_n)$ *are extended in all ways.*
  (P2) For a finite fixed $X \subseteq V(G_n)$, a new node $z_X \notin V(G_n)$ is added whose neighbours in $V(G_n)$ are exactly the nodes of X.

The graph G is then $\lim_{n\rightarrow\infty} G_n$. We refer to the graphs G_n as *time-steps* in the evolution of G. If (P2) is never used at any time-step, then we say that the corresponding construction sequence is *pure*; otherwise, the constructing sequence is *mixed*. A graph G is *pure* if it has a pure constructing sequence. Otherwise, we say that G is *mixed*. In a mixed constructing sequence, the nodes added in (P2) are called *extra* nodes. We note that R has both an (A)- and (B)-constructing sequence where (P2) is used at each time-step. (We leave the details as an exercise.)

We note that a graph G formed by an (A)-constructing sequence has property (A). The converse also holds.

Lemma 1. *Let G be a graph with $V(G) = \omega$, and fix a finite induced subgraph H. If $\mathcal{P}$ is A or B, and if G has property ($\mathcal{P}$), then G has an ($\mathcal{P}$)-constructing sequence $(G_n : n \in \omega)$ with $G_0 = H$.*

Proof. Assume that $\mathcal{P}$=A. By relabelling nodes if necessary, we may assume that $0 \in V(G_0)$. We will show how to construct a chain of graphs $(G_n : n \in \omega)$ with the following property: for each $n \in \omega$, $n \in V(G_n)$ and G_n is a finite induced subgraph of G. Note that from this property, it follows that $\lim_{n\rightarrow\infty} G_n = G$.

Let $G_0 = H$. Inductively, assume that G_n is defined and finite. For each node y of G_n, and each subset $X \subseteq N[y]$, let $Y = V(G_n) \setminus X$. Since G has property (A), there exists a node $z_X \in V(G) \setminus V(G_n)$ that is joined to all nodes in X, and none in Y. Let V' be the set of such nodes z_X, exactly one for each subset $X \subseteq N[y]$ in G_n. Define G_{n+1} to be the finite subgraph of G induced by $V(G_n) \cup V' \cup \{n+1\}$. It is clear that, for all $n \in \omega, G_n \leq G_{n+1}$, so $(G_n : n \in \omega)$ is a chain. Adding each node z_X in V' to G_n corresponds to an application of process (P1). If the node $n+1$ is not in V', then adding $n+1$ to G_n corresponds to one application of process (P2).

The proof for property (B) is similar, and so is omitted. □

Corollary 1. *If $\mathcal{P}$ is either A or B, then the following are equivalent.*

(1) *The graph G has property ($\mathcal{P}$).*
(2) *The graph G has an ($\mathcal{P}$)-constructing sequence.*

3. Many models

Our main results concern graphs with properties (A) and (B). We first show that properties (A) and (B) are not $\aleph_0$-*categorical*; that is, there are many non-isomorphic graphs that satisfy these properties.

A *homomorphism* from G to H is a mapping $f : V(G) \to V(H)$ that *preserves edges*; more precisely, $xy \in E(G)$ implies that $f(x)f(y) \in E(H)$. We usually write $f : G \to H$ or just $G \to H$. If $G \to H$ and $H \to G$, then we say that G and H are *homomorphically equivalent*, and write $G \leftrightarrow H$. See [17] for more on graph homomorphisms.

Theorem 3. *Let H be a finite graph. Let G be an infinite pure (B) graph with a pure (B)-constructing sequence $(G_n : n \in \omega)$, where $G_0 = H$. Then $H \leftrightarrow G$.*

Proof. As $H \leq G$, we have that $H \to G$. To show that $G \to H$, we construct a homomorphism by induction on $n \in \omega$. Let f_0 be the identity map on $G_0 = H$. Suppose that $f_n : G_n \to H$ is a homomorphism extending f_0. Let $z \in V(G_{n+1}) \backslash V(G_n)$. Then by the definition of (P1) there exist a node $y \in V(G_n)$ and a subset X of $N(y)$, so that z is only joined to nodes of X. So in G_{n+1}, $N(z) \subseteq N(y)$. We label this node y as y_z. Since f is a homomorphism, $f_n(y_z) \neq f_n(x)$ for all $x \in N(y_z)$. Hence, we may map z to $f_n(y_z)$ and preserve edges. Therefore, the map $f_{n+1} : G_{n+1} \to H$ defined by

$$f_{n+1}(z) = \begin{cases} f_n(z) & \text{if } z \in V(G_n); \\ f_n(y_z) & \text{else,} \end{cases}$$

is a homomorphism. The map $F : G \to H$ defined by $\bigcup_{n\in\omega} f_n$ is a homomorphism. □

The following Corollary is immediate from Theorem 3.

Corollary 2. *For a fixed finite graph H, let $G(H)$ be an infinite pure (B) graph with a pure (B)-constructing sequence $(G_n : n \in \omega)$ such that*

$G_0 = H$. Then the following hold:

(1) *$\chi(G(H)) = \chi(H)$ and $\omega(G(H)) = \omega(H)$.*
(2) *If H and H' are not homomorphically equivalent, then $G(H) \ncong G(H')$.*

We note that there are infinite families of non-homomorphically equivalent finite graphs; see [5]. Hence, there are at least $\aleph_0$ many non-isomorphic pure (B) graphs; see Theorem 6. This contrasts with the situation for pure (A) graphs.

Theorem 4. *There is a unique pure (A) graph, up to isomorphism.*

We will defer the proof of Theorem 4 to Section 4, since our proof will make heavy use of the inexhaustibility property which is discussed there. If n is a positive integer, then a graph G is *n-existentially closed* or *n-e.c.* if each n-subset S of $V(G)$ can be extended in all ways. Hence, a graph G is e.c. if it is n-e.c. for all positive integers n. It is well-known that for every constant $p \in (0, 1)$, and fixed n a positive integer, almost all finite random graphs with edges chosen independently with probability p are n-e.c. We use this property to give the maximum possible cardinality of non-isomorphic mixed graphs with property (A).

Theorem 5. *There are $2^{\aleph_0}$ many non-isomorphic infinite mixed graphs with property (A).*

Proof. Fix $n \geq 5$. Let $G_0 = C_{n+1}$, the chordless cycle on $n + 1$ nodes. Assume that G_i is defined and finite. To form G_{i+1}, first apply process (P1) and extend all subsets of closed neighbour sets of G_i to form G'_{i+1}. Then apply process (P2) a finite number of times by extending all n-subsets of nodes of G'_{i+1} to form G_{i+1}. Define $G(n) = \lim_{n\to\infty} G_i$. The graph $G(n)$ has Property (A) with $(G_i : i \in \omega)$ an (A)-constructing sequence, and is clearly n-e.c. Note that $G(n)$ is mixed, since (P2) is used in the constructing sequence. However, there is no node in $G(n)$ that is joined to *each node of* G_0, so $G(n)$ is not $(n+1)$-e.c. To see this, we proceed by induction on i. Assume that there is no node in G_i joined to all of G_0. In G_{i+1}, the nodes that are added to G_i are of two types: 1) extending subsets of closed-neighbour sets in G_i, or 2) extending arbitrary n-subsets in G_i. The nodes of type 2 can never be joined to all the $n + 1$ nodes of G_0. Now consider nodes of type 1. Assume, to obtain a contradiction, that $V(G_0) \subseteq N[y]$ for some y in G_i. Then y cannot equal an element of G_0, as G_0 contains no universal nodes. Hence, $V(G_0) \subseteq N(y)$ which contradicts our induction hypothesis. Therefore, there is no type 1 node in $V(G_{i+1})\backslash V(G_i)$ joined to each node of G_0.

Since any n-e.c. graph, where $n \geq 2$, is connected, it follows that each graph $G(n)$ is connected. Now, let X be an infinite subset of ω, listed as $X = \{n_i : i \in \omega\}$. Define

$$G(X) = \biguplus_{i \in \omega} G(n_i).$$

Hence, the connected components of $G(X)$ are the $G(n_i)$. Then $G(X)$ satisfies (A), since property (A) is preserved by taking disjoint unions, as is readily verified. Let Y be an infinite subset of ω with $X \neq Y$. Let $n \in X \backslash Y$. Then $G(X)$ contains a connected component that is n-e.c. but not $(n+1)$-e.c. However, there is no such connected component in $G(Y)$; thus, $G(X) \ncong G(Y)$. As there are $2^{\aleph_0}$ many infinite subsets of ω, there are $2^{\aleph_0}$ many non-isomorphic (A) graphs. As there is a unique isomorphism type of pure (A) graph by Theorem 4, there are $2^{\aleph_0}$ many non-isomorphic mixed (A) graphs. □

Theorem 6. *There are $2^{\aleph_0}$ many non-isomorphic infinite mixed graphs with property (B) but not (A). There are exactly $\aleph_0$ many non-isomorphic infinite pure graphs with property (B) but not (A).*

Proof. Let G be a pure (B) graph, with a (B)-constructing sequence $(G_i : i \in \omega)$ so that $G_0 = K_n$, for a fixed $n \geq 4$. Since G is pure, G_{i+1} is obtained from G_i only by process (P1), for each positive integer i. By Corollary 2, $\chi(G) = \chi(G(K_n)) = n$.

Claim: $G = G' \uplus \overline{K_{\aleph_0}}$, where G' is a connected graph with $\chi(G') = n$.

To see this, note that G_{i+1} was constructed by adding nodes joined to some or no nodes of G_i. Suppose that $G_i = G_1(i) \uplus G_2(i)$, where $(G_1(i))$ is connected and $G_2(i)$ is independent. Let $G_1(i+1)$ be those nodes in G_{i+1} that are joined to *some* node of G_i, and let $G_2(i+1)$ be the nodes in G_{i+1} that are not joined to *any* node of G_i. By definition of (P1), the new nodes of G_{i+1} are either joined to the neighbourhood of a node in G_i, and thus must be connected to $G_1(i)$, or they are independent, and hence they are part of $G_2(i+1)$. Therefore, $(G_1(i+1))$ is connected and contains $G_1(i)$, and $G_2(i+1)$ forms an independent set in G_{i+1} and contains $G_2(i)$. Moreover, $G_2(i+1)$ properly contains $G_2(i)$, because according to process (P1), for every node z of G_i, for the choice $X = \emptyset$, a new independent node z_X is added. Note that $G_{i+1} = G_1(i+1) \uplus G_2(i+1)$. Note also that $G_0 = K_n \uplus H_0$, where H_0 is the empty graph. Let $\lim_{i\to\infty}(G_1(i)) = G'$ and let $\lim_{i\to\infty}(G_2(i)) = H$. The graph G' is connected, as each graph $G_1(i)$ is connected, and H is independent, since each $G_2(i)$ is independent. Since the cardinality of

$G_2(i)$ is strictly increasing, $H \cong \overline{K_{\aleph_0}}$. Also, since $G(i) = G_1(i) \uplus G_2(i)$ for each i, and $\lim_{i\to\infty} G_i = G$, $G = G' \uplus H$. As $G_0 \leq G' \leq G$, it is immediate that G' has chromatic number n.

Now let $\Omega = \{K_n : n \geq 4\}$. For a fixed infinite $X \subseteq \Omega$, define a graph $G(X)$ as follows. Let $X = \{K_{n_i} : i \in \omega\}$. As in Corollary 2, let $G(K_{n_i})$ be the graph obtained from $G_0 = K_{n_i}$ by a pure (B)-constructing sequence. By setting $G(K_{n_i})$ to be the graph G defined earlier in the proof, we obtain that $G(K_{n_i}) = G'(n_i) \uplus \overline{K_{\aleph_0}}$, where $G'(n_i)$ is a connected graph with $\chi(G'(i)) = n_i$. Define $G(X)$ as $\biguplus_{i\in\omega} G(K_{n_i})$. Then $G(X)$ has property (B), but note that $G(X)$ cannot have (A) by Theorem 1, since the chromatic number of each connected component is finite. Now if X, Y are infinite subsets of Ω with $X \neq Y$, then suppose that $K_n \in X \backslash Y$. By the Claim, there is no component in $G(Y)$ with chromatic number n, so $G(X) \not\cong G(Y)$. Hence, there are $2^{\aleph_0}$ many non-isomorphic (B) graphs.

Let G be a pure (B) graph with pure (B)-constructing sequence $(G_i : i \in \omega)$. It is not hard to see that G is determined up to isomorphism by the finite graph G_0. As there are only $\aleph_0$ many non-isomorphic choices for G_0, there are at most $\aleph_0$ non-isomorphic pure (B) graphs. By Corollary 2 there are at least $\aleph_0$ non-isomorphic pure (B) graphs, at least one with chromatic number n, for each $n \in \omega$. Hence, there are exactly $\aleph_0$ many non-isomorphic pure (B) graphs, and therefore, by the last sentence in the previous paragraph, there are $2^{\aleph_0}$ many non-isomorphic mixed (B) graphs. □

We note that the infinite random graph R has property (A) (and therefore (B)), but has neither a pure (A)- nor (B)-constructing sequence. To see this in the case of property (A), we note that any pure (A) graph is disconnected. Let G be a pure (A) graph with pure (A)-constructing sequence $(G_i : i \in \omega)$. At each time step G_n, where $n > 1$, at least two isolated nodes are introduced. For a fixed $n > 1$, call two such nodes in G_n u and v. An inductive argument shows that in the following time-steps G_r, with $r > n$, u and v remain in different components of G_r. Hence, there are at least two connected components in G. However, since R is e.c., it is connected of diameter 2. Therefore, R cannot have a pure (A)-constructing sequence. We proved in Corollary 2 that any pure (B) graph has finite chromatic number, and so R cannot have a pure (B)-constructing sequence.

4. Fractal and other properties

A graph G is *inexhaustible* if for all $x \in V(G)$, we have that $G - x \cong G$. The graph R is inexhaustible, as are the infinite complete and null graphs. For more on inexhaustible graphs, the reader is directed to [6] and [15]. We prove that the same property holds for graphs satisfying properties (A) or (B). Inexhaustibility is an example of a *fractal property* of graphs: an inexhaustible graph is resilient under node deletion, a property observed in the actual web graph in [11]. For more on fractal properties of graphs, the reader is directed to [7].

Theorem 7. *If G is a fixed graph with property (B), then G is inexhaustible.*

Proof. Let $(G_n : n \in \omega)$ be a (B)-constructing sequence for G. If n is a positive integer, then a set of nodes in G_n is called *n-special* if it includes nodes of $V(G_n)\backslash V(G_{n-1})$. We introduce the following notation for subsets of $V(G)$. Let S_1 be the set of nodes added to G_0 in time step 1 by extending sets of nodes of G_0. In general, in G_r, let S_r be the set of nodes extending G_{r-1}. Let $S_{1,1} = S_1$. In S_2, there are nodes $S_{2,1} \subseteq S_2$ extending G_0 as S_1 does, and nodes $S_{2,2} \subseteq S_2$ extending 1-special sets of nodes. Note that $S_{2,1} \cup S_{2,2} = S_2$ and $S_{2,1} \cap S_{2,2} = \emptyset$. In general, in S_r we define a finite sequence $(S_{r,i} : 1 \leq i \leq r)$ of sets of nodes partitioning S_r, with each $S_{r,i}$ consisting of the nodes that extend the $(i-1)$-special sets of G_{r-1}. In particular, $S_{r,r}$ is the *only* set extending $(r-1)$-special sets of nodes. If $1 \leq i \leq r-1$, then the nodes in the set $S_{r,i}$ extend the same subsets that $S_{r-1,i}$ does.

Let $G_{\infty,0} = G_0$. To define $G_{\infty,1}$, we form disjoint sets of nodes $(S_{\infty,1,i} : i \in \omega)$, each disjoint from $V(G_{\infty,0})$ and of the same cardinality as $S_{1,1}$, and let

$$V(G_{\infty,1}) = V(G_{\infty,0}) \cup \bigcup_{i \in \omega} S_{\infty,1,i}.$$

Now let each $S_{\infty,1,i}$ extend G_0 as $S_{1,1}$ does. More precisely, the subgraph of $G_{\infty,1}$ induced by $V(G_0)$ and $S_{\infty,1,i}$ is isomorphic to G_1. Moreover, $\bigcup_{i \in \omega} S_{\infty,1,i}$ is an independent set in $G_{\infty,1}$. It follows from the definition that $G_{\infty,1}$ contains infinitely many subgraphs isomorphic to G_1.

Assume that $G_{\infty,r}$ is defined, countable, and contains infinitely many subgraphs isomorphic to G_r. To define $G_{\infty,r+1}$, form disjoint sets of nodes $(S_{\infty,r,i} : i \geq r)$, each disjoint from $V(G_{\infty,0})$ and of the same cardinality as $S_{r,r}$. Let

$$V(G_{\infty,r}) = V(G_{\infty,r-1}) \cup \bigcup_{i \geq r} S_{\infty,r,i}.$$

Let each of the $S_{\infty,r,i}$ extend one of the subgraphs of $G_{\infty,r}$ isomorphic to G_r as $S_{r,r}$ does, and let $\bigcup_{i \geq r} S_{\infty,r,i}$ form an independent set in $G_{\infty,r+1}$.

Clearly, $G_{\infty,r+1}$ contains infinitely many subgraphs isomorphic to G_r. See Figure 1.

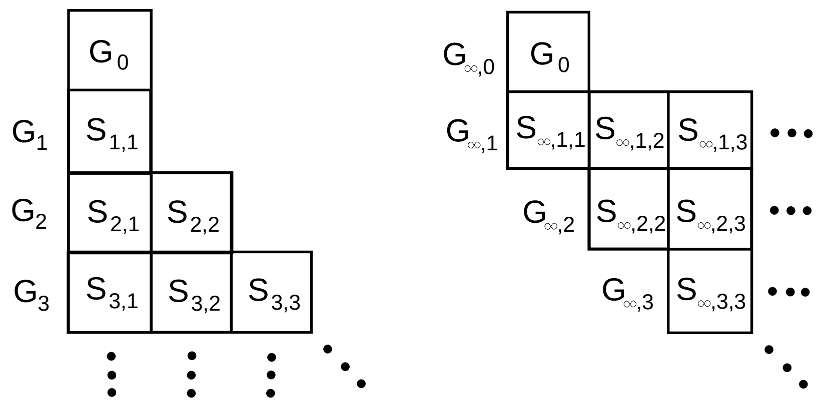

FIGURE 1. The graphs G and G_∞.

Claim 1: $G \cong G_\infty$.

We define a mapping $f : G \to G_\infty$ as follows. The map f sends G_0 to $G_{\infty,0}$ via the identity map. For each $i, j \in \omega\backslash\{0\}$, map $S_{i,j}$ isomorphically onto $S_{\infty,j,i}$. (Hence, "columns" in G are mapped to "rows" in G_∞; see Figure 1.) It is straightforward to see that f is an isomorphism. Claim 1 follows.

Now, we prove that G_∞ is inexhaustible. Fix $x \in V(G_\infty)$.

Case 1. $x \notin V(G_{\infty,0})$.

Let m be the largest non-negative integer so that $G_{\infty,m}$ does not contain x. Let $f_m : G_{\infty,m} \to G_{\infty,m}$ be the identity mapping.

Claim 2: For all $r > m$, $G_{\infty,r} \cong G_{\infty,r} - x$ via an isomorphism f_r extending f_{r-1}.

We proceed by induction on r. We use the *back-and-forth method* to define f_r, which is a two player game of perfect information played in countably many steps on two graphs X_0, X_1. The players are named the *duplicator* and the *spoiler*. (The names come from the facts that the duplicator is trying to show the graphs are alike, while the spoiler is trying to show they are different.) A move consists of a choice of a node from either graph, and the spoiler makes the first move. The players take turns choosing nodes from the $V(X_i)$, so that if one player chooses a node from $V(X_i)$, the other must choose a node of $V(X_{i+1})$ (the indices are mod 2). The game begins in our case with a fixed isomorphism f between two induced subgraphs Y_0 and Y_1 of X_0 and X_1, respectively. Players cannot choose previously chosen nodes, or nodes in a Y_i. After n rounds, this gives rise to a list of nodes $Y_0 \cup \{a_i : 1 \leq i \leq n\}$ from X_0 and $Y_1 \cup \{b_i : 1 \leq i \leq n\}$ from X_1. The duplicator wins if for every $n \geq 1$, the subgraph induced by $Y_0 \cup \{a_i : 1 \leq i \leq n\}$ is isomorphic to the subgraph induced by $Y_1 \cup \{b_i : 1 \leq i \leq n\}$ via an isomorphism extending f mapping a_i to b_i, for every $1 \leq i \leq n$. Otherwise, the spoiler wins. From this it follows that the duplicator has a winning strategy if and only if X_0 and X_1 are isomorphic via an isomorphism extending f. See [8] for more on the back-and-forth method.

We let

$$X_0 = G_{\infty,r}, X_1 = G_{\infty,r} - x, f = f_{r-1}, Y_0 = G_{\infty,r-1}, Y_1 = G_{\infty,r-1} - x.$$

Going forward, suppose that the spoiler chooses y in $G_{\infty,r}$, where y is not in $G_{\infty,r-1}$. We will assume that y is a node added at time-step G_r by process (P1) (the argument for process (P2) is similar and so is omitted). Then y extends some set A in $N(z)$, for some A and z in $G_{\infty,r-1}$. Let $A' = f_{r-1}(A)$, and $z' = f_{r-1}(z)$ in $G_{\infty,r-1} - x$. Hence, there is a y' in $G_{\infty,r}$ extending A' as y extends A. If $y' \neq x$, then the duplicator chooses this node. If $y' = x$, then the duplicator may choose any of the infinitely many nodes in $G_{\infty,r} - x$ that also extends A' as y' does. Going back is similar. Since any two nodes in $V(G_{\infty,r}) \backslash V(G_{\infty,r-1})$ are non-joined, Claim 2 follows.

The map $\bigcup_{r\in\omega} f_r : G_\infty \to G_\infty - x$ is an isomorphism.

Case 2. $x \in V(G_{\infty,0})$.

Given the finite graph $G_{\infty,0}$, define the infinite graph $G'_{\infty,0}$ as follows. For each node y of $G_{\infty,0}$, add infinitely many new pairwise non-joined nodes y_i with the property that y_i has the same neighbours in $G_{\infty,0}$ as y does. (We think of the y_i as nodes extending $N(y)$. Hence, $G'_{\infty,0} \leq G_{\infty,1}$ in G_∞.) It is straightforward to see that $G'_{\infty,0}$ is inexhaustible. (See Figure 2 for a depiction of the graph C'_4.)

Define a new graph G'_∞ that is constructed as G_∞ was, but beginning with $G'_{\infty,0}$, rather than $G_{\infty,0}$.

Claim 3: $G'_\infty \cong G_\infty$.

To prove Claim 3, we first prove that $G'_{\infty,1}$ and $G_{\infty,1}$ are isomorphic by extending the identity mapping g_0 between $G_{\infty,0} \leq G'_{\infty,1}$ and $G_{\infty,0} \leq G_{\infty,1}$. Suppose that the spoiler chooses a node y' in $V(G'_{\infty,1}) \setminus V(G_{\infty,0})$. Consider the case when y' was added by process (P1). (The argument for the case when y' is added by process (P2) is similar, and so is omitted.) The duplicator can respond with a node $y \in V(G_{\infty,1})$ joined to $A \subseteq V(G_{\infty,0})$. Going back is similar. Since no two nodes in $V(G_{\infty,1}) \backslash V(G_{\infty,0})$ or in $V(G'_{\infty,1}) \backslash V(G_{\infty,0})$ are joined, as

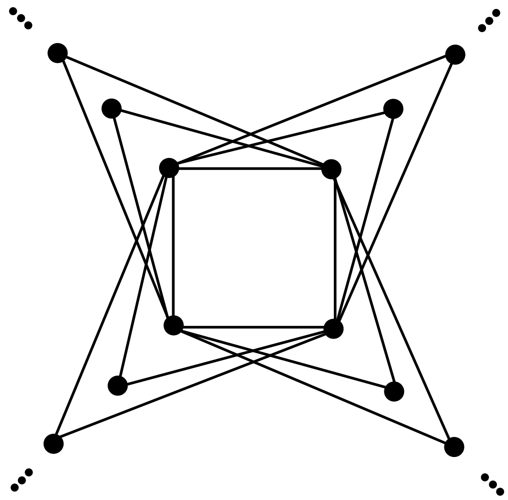

FIGURE 2. The graph C'_4.

we noted in Case 1, the duplicator can win. Using similar arguments, we obtain, for each $r \in \omega$, isomorphisms $g_r : G'_{\infty,r} \to G_{\infty,r}$ so that if $r \geq 1$, then $g_r \restriction G'_{\infty,r-1} = g_{r-1}$. The map

$$\bigcup_{r\in\omega} g_r : G'_\infty \to G_\infty$$

is an isomorphism.

To finish Case 2, it is therefore sufficient to show that G'_∞ is inexhaustible. Choose $y \in V(G'_\infty)$. If $y \in V(G'_{\infty,0})$, since $G'_{\infty,0}$ is inexhaustible, there is an isomorphism $g_0 : G'_{\infty,0} \to G'_{\infty,0} - y$. As in Case 1, extend g_0 to isomorphisms $g_r : G'_{\infty,r} \to G'_{\infty,r} - y$, for all $r \geq 0$, by back-and-forth. The map

$$\bigcup_{r\in\omega} g_r : G'{}_\infty \to G'{}_\infty - x$$

is an isomorphism. If $y \notin V(G'_{\infty,0})$, then proceed as in Case 1. $\square$

Recall that a graph which has an (A)-constructing sequence is *pure* if at each time step, process (P1) is used; it is *mixed* otherwise. One may ask whether there are many non-isomorphic pure (A) graphs. Using Theorem 7, we prove that the answer is negative, and thereby prove Theorem 4 of Section 3 above. This is in stark contrast to the situation for mixed (A) graphs, as proved in Theorem 5. If G_0 is a finite graph, we use the notation $\uparrow G_0$ for the unique (up to isomorphism) graph that results by applying the (P1) process for property (A) recursively to G_0. It follows that every pure (A) graph is of the form $\uparrow G_0$ for some finite graph G_0. It is not hard to see that $\uparrow (G \uplus H) \cong \uparrow G \uplus \uparrow H$.

Proof of Theorem 4: It is enough to show that if G_0 and H_0 are finite graphs, then $\uparrow G_0 \cong \uparrow H_0$. For this let $(G_i : i \in \omega)$ and $(H_i : i \in \omega)$ be (A)-constructing sequences for $G = \uparrow G_0$ and $H = \uparrow H_0$, respectively. As H is $\aleph_0$-universal by Theorem 1, $G \leq H$. In particular, there is some $n \in \omega$ so that $G_0 \leq H_n$. Delete from H all the finitely many nodes in $S = V(H_n) \setminus V(G_0)$. At time-step $n+1$, we are left with a copy of G_1 extending G_0 (as it is extended in G), and finitely many isolated nodes, say m of them (that were either joined to no node of H_n in H, or were joined only to nodes that we deleted). Hence, $H_{n+1} - S \cong G_1 \uplus \overline{K_m}$. Since G_1 extends G_0, we have that $H - S \cong \uparrow (G_0 \uplus \overline{K_m}) \cong \uparrow G_0 \uplus \uparrow \overline{K_m}$.

At each time-step r in the construction of G $(r > 0)$, nodes with no edges to G_{r-1} are added to G_r. These nodes give rise to connected components of G of the form $\uparrow K_1$. Hence, G contains infinitely many connected components of the form $\uparrow K_1$. It follows that G is isomorphic to the graph J, which consists of the disjoint union of $\uparrow G_0$ and infinitely many disjoint copies of $\uparrow K_1$.

As H is inexhaustible by Theorem 7, it follow that $H - S \cong H$. But then

$$H \cong H - S \cong \uparrow (G_0 \uplus \overline{K_m}) \cong \uparrow G_0 \uplus \uparrow \overline{K_m} \cong J \cong G,$$

since $\uparrow \overline{K_{\aleph_0}} \uplus \uparrow \overline{K_m} \cong \uparrow \overline{K_{\aleph_0}}$. □

The unique isomorphism type of Theorem 4 we name R_N, since it is locally isomorphic to R. We do not know much about R_N. The infinite random graph R is *indivisible*: whenever the nodes of R are coloured red or blue, there is a monochromatic induced subgraph isomorphic to R. For more on indivisible graphs, see [12, 15]. A graph without this property is *divisible*. A graph G so that $R \leq G$ is necessarily indivisible, since R is itself indivisible. Therefore, by Theorem 1, a graph with property (A), such as R_N, is indivisible. It is not hard to see that a graph with at least one edge and with finite chromatic number is divisible, so by the proof of Theorem 5 for (B), there are examples of graphs with (B) that are divisible.

A *ray* is an infinite path that extends indefinitely in one direction; a *double ray* is an infinite path that extends indefinitely in two directions. A *one-way Hamilton path* is a spanning subgraph that is a ray, while a *two-way Hamilton path* is a spanning subgraph that is a double ray. The graph R contains one- and two-way Hamilton paths.

Theorem 8. *If G has property (B), then the connected components of G have one and two-way Hamilton paths. In particular, G has a 1-factor.*

Proof. Let G' be a fixed connected component of G. We prove that G' has a one-way Hamilton path; the existence of a two-way Hamilton path is similar. Without loss of generality, let $V(G') = \omega$.

Define P_0 be the subgraph induced by $\{0\}$. Assume that there is a path $P(n)$ in G' containing the nodes $\{0, \ldots, n\}$, and that the nodes of $P(n)$ are $x_1, \ldots, x_s$. If the node $n+1$ equals some x_i, then let $P(n+1) = P(n)$. Otherwise, assume that $n+1$ is not a node in $P(n)$. As G' is connected, the node $n+1$ is connected by a path Q to some node x_i of $P(n)$. Let Q be the path $y_1 y_2 \cdots y_{t-1} y_t$, where $y_1 = n+1$ and $y_t = x_i$. As x_{i+1} and y_{t-1} are in $N(x_i)$, by (B) there is a node z_1 in G, and hence, in G', that is joined to both x_{i+1} and y_{t-1}, and is not a node of $P(n)$ nor Q.

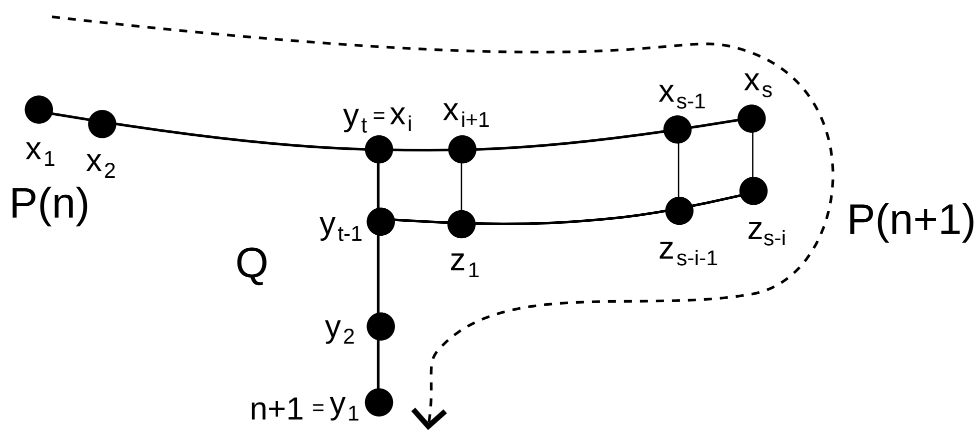

FIGURE 3. A Hamilton path containing $\{0, \ldots, n+1\}$.

Proceeding inductively, we obtain a path of nodes $z_1 z_2 \cdots z_{s-i}$ in G' so that z_{j+1} is joined to both z_j and x_{i+j+1}. Then the path $P(n)$ followed by the path $z_{s-i} \cdots z_2 z_1$ is a path $P(n+1)$ containing $\{0, \ldots, n+1\}$. The desired Hamilton path of G' is $\lim_{n\to\infty} P(n)$. See Figure 3. □

A *bipartite core* is a graph on $i + j$ nodes that contains at least one bipartite clique $K_{i,j}$ as a subgraph. Bipartite cores arise in the web graph as representing communities of users who have pages on some common topic; see [19, 20]. The findings of [18] reveal that the web graph contains at least several hundred thousand such cores. Many bipartite cliques almost surely arise in the evolving copying model of

[19], so we should expect that graphs satisfying (B) would contain many bipartite cliques. The infinite random graph R contains every countable bipartite graph as an induced subgraph, so there are many such bipartite cores in graphs with the e.c. property, or with (A) by Theorem 1. The following theorem proves a similar result for graphs with (B). The infinite bipartite graph with each vertex class infinite is written $K_{\aleph_0,\aleph_0}$.

Theorem 9. *If G is a graph with (B) so that G_0 has at least one edge, then $K_{\aleph_0,\aleph_0} \leq G$. In particular, there are infinitely many node-disjoint bipartite cores in a graph satisfying (B).*

Proof. Let S_1 be some edge of G (by the fact that G_0 has edges, we know that G also has edges). Assume that $S_n \leq G$ is isomorphic to $K_{n,n}$ with $S_1 \leq S_n$. Suppose that S_n has nodes $A \cup B$, with $A = \{x_i : 1 \leq i \leq n\}$, $B = \{y_i : 1 \leq i \leq n\}$ consisting of independent sets. By property (B), there is a node x_{n+1} not in $A \cup B$ that is joined to each node of B. Since $A \cup \{x_{n+1}\} \subseteq N(y_1)$, by (B), there is a node y_{n+1} not in $A \cup B \cup \{x_{n+1}\}$ that is joined to each node of $A \cup \{x_{n+1}\}$. Then the graph S_{n+1} consisting of S_n along with x_{n+1} and y_{n+1} is a bipartite clique strictly containing S_n. The graph $H = \bigcup_{n \geq 1} S_n \leq G$ is isomorphic to $K_{\aleph_0,\aleph_0}$. □

References

[1] W. Aiello, F. Chung, L. Lu, A random graph model for power law graphs, *Proc. ACM Symp. on Theory of Computing*, (2000) 171-180.

[2] W. Aiello, F. Chung, L. Lu, Random evolution in massive graphs, 42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001), 510–519, IEEE Computer Soc., Los Alamitos, CA, 2001.

[3] A. Barabási, R. Albert, Emergence of scaling in random networks, *Science* **286** (1999) 509-512.

[4] A. Barábsi, R. Erzsébet, T. Vicsek, Deterministic scale-free networks, *Physica A* **299** (2001) 559-564.

[5] A. Bonato, A family of universal pseudo-homogeneous G-colourable graphs, *Discrete Math.* **247** (2002) no. 1-3, 13-23.

[6] A. Bonato, D. Delić, On a problem of Cameron's on inexhaustible graphs, to appear in *Combinatorica*.

[7] A. Bonato, P. Cameron, D. Delić, and S. Thomassé, Generalized pigeonhole properties of graphs and oriented graphs, *Europ. J. Combin.* **23** No. 3. 257-274.

[8] P.J. Cameron, Oligomorphic groups and homogeneous graphs, in: *Graph symmetry* (Montreal, PQ, 1996), 23–74, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 497, Kluwer Acad. Publ., Dordrecht, 1997.

[9] P.J. Cameron, The random graph, in: *Algorithms and Combinatorics* **14** (R.L. Graham and J. Nešetřil, eds.), Springer Verlag, New York (1997) 333-351.

[10] C. Cooper, A. Frieze, On a general model of web graphs, *Proceedings of ESA* (2001) 500-511.

[11] S. Dill, R. Kumar, K. McCurley, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, Self-similarity in the web, in: *Proceedings of International Conference on Very Large Data Bases*, 69–78, Rome, 2001.

[12] M. El-Zahar, N. Sauer, The indivisibility of the homogeneous K_n-free graphs, *J. Combin. Theory Ser. B* **47** (1989), no. 2, 162-170.

[13] P. Erdős, A. Rényi, Asymmetric graphs, *Acta Math. Acad. Sci. Hungar.* **14** (1963) 295–315.

[14] P. Erdős, A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960) 17-61.

[15] R. Fraïssé, *Theory of Relations*, Revised edition, with an appendix by Norbert Sauer, (North Holland, Amsterdam, 2000).

[16] R. Fraïssé, Sur certains relations qui généralisent l'ordre des nombres rationnels, *C.R. Acad. Sci. Paris* **237** (1953) 540-542.

[17] G. Hahn, C. Tardif, Graph homomorphisms: structure and symmetry, in: G. Hahn and G. Sabidussi, eds. *Graph symmetry*, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **497** (Kluwer Acad. Publ., Dordrecht, 1997) 107–166.

[18] R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, Trawling emerging cyber-communities automatically, *Proc. 8th WWW Conf.*, 1999.

[19] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, Stochastic models for the Web graph, *Proceedings of the 41th IEEE Symp. on Foundations of Computer Science*, pages 57-65, 2000.

[20] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, The web as a graph, Proc. 19th ACM SIGACT-SIGMOD-AIGART Symp. Principles of Database Systems, May 15 -17, 2000, Dallas, Texas, USA, 1–10. ACM, 2000.

[21] A.H. Lachlan, R.E. Woodrow, Countable ultrahomogeneous undirected graphs, *Trans. Amer. Math. Soc.* **262** (1980) no. 1, 51-94.

Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada, N2L 3C5

E-mail address: `abonato@wlu.ca`

Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, Canada, B3H 3J5

E-mail address: `janssen@mathstat.dal.ca`