EuroComb 2005 DMTCS proc. **AE**, 2005, 223–228

Infinite limits of the duplication model and graph folding

Anthony Bonato[1†] and Jeannette Janssen[2‡]

[1] *Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada, N2L 3C5*
[2] *Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, Canada, B3H 3J5*

We study infinite limits of graphs generated by the duplication model for biological networks. We prove that with probability 1, the sole nontrivial connected component of the limits is unique up to isomorphism. We describe certain infinite deterministic graphs which arise naturally from the model. We characterize the isomorphism type and induced subgraph structure of these infinite graphs using the notion of dismantlability from the theory of vertex pursuit games, and graph homomorphisms.

Keywords: massive networks, duplication model, infinite random graph, folding, adjacency property, graph homomorphism

Much recent attention has focused on stochastic models of real-world massive self-organizing networks. The reader is directed to the two recent surveys [1, 2]. In self-organizing networks, each node acts as an independent agent, which will base its decision on how to link to the existing network on local knowledge. As a result, the neighbourhood of a new node will often be similar to that of an existing node. An important example of such a network consists of the protein-protein interactions in a living cell. The *duplication model* of [7] was designed to model such biological networks. The two parameters of the model are a real number $p \in (0, 1)$, and a finite undirected graph H. New nodes are introduced over a countable sequence of discrete time-steps, with the graph at G_0 at time $t = 0$ equalling H. At time $t + 1$, a node u is chosen uniformly at random from the existing nodes in G_t. To form G_{t+1}, a new node v_{t+1} is added. For each of the neighbours z of u, we add the edge zv_{t+1} to the edges of G_{t+1} with probability p.

In this paper, we study the infinite graphs that result when time is allowed to go to infinity. The study of infinite limits of random graph models can give new insight into the properties of the model. If $p \in (0, 1)$, then the classical results of Erdős, Rényi [8] imply that with probability 1, an infinite limit of $G(n, p)$ graphs is isomorphic to a unique isomorphism type of graph written R. This well-known result may appear contradictory at first, since it suggests an infinite random process with a deterministic conclusion. The deterministic graph R is often called the *infinite random graph*, and is the unique isomorphism type of countable graph satisfying the *existentially closed* or *e.c.* adjacency property: for all finite disjoint sets of nodes X and Y, there is a node not in $X \cup Y$ that is joined to all of X and none of Y. A logically

†Supported by an NSERC Discovery grant, and from a MITACS grant.
‡Supported by an NSERC Discovery grant, and from a MITACS grant.

weaker adjacency property is *locally e.c.,* introduced by the authors in [3]. (In [3], locally e.c. is referred to as property (B).) If y is a node of G, then $N_G(y) = \{z \in V(G) : yz \in E(G)\}$ is the *neighbour set* of y in G. A graph G is *locally e.c.* 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 not in $\{y\} \cup X \cup Y$ that is joined to X and not to Y. The locally e.c. property is therefore a variant of the e.c. property that applies only to sets contained in the neighbour set of a node.

A preliminary connection between the locally e.c. property and the duplication model was made explicit by the following theorem, first proven in [3]. If $(G_t : t \in \mathbb{N})$ is a sequence of graphs with G_t an induced subgraph of G_{t+1}, then define the *limit* of the G_t, written $G = \lim_{t\to\infty} G_t$, by $V(G) = \bigcup_{t\in\mathbb{N}} V(G_t), E(G) = \bigcup_{t\in\mathbb{N}} E(G_t)$.

Theorem 1 *Fix $p \in (0,1)$, and $G_0 = H$ a finite graph. With probability 1, a limit $G = \lim_{t\to\infty} G_t$ of graphs generated by the duplication model is locally e.c.*

As was proved in [3], in stark contrast to the e.c. property, there are uncountably many non-isomorphic locally e.c. graphs. The main goal of the present article is to introduce countably infinite graphs, which are, with high probability, the *unique* limits of the duplication model, up to the presence of isolated nodes. These graphs are therefore analogues of the infinite random graph for the duplication model.

Fix H a finite graph. Let $R_0 \cong H$. Assume that R_t is defined and is finite. For each node $y \in V(R_t)$, and each subset $X \subseteq N(y)$, add a new node $z_{y,X}$ joined only to X. This gives the graph R_{t+1} which contains R_t as an induced subgraph. Define $R_H = \lim_{t\to\infty} R_t$. Observe that the (deterministic) graph R_H is locally e.c.

Suppose that H, J are finite graphs, with $v \in V(J)$. Define $H \preccurlyeq_v J$ if there is a node u in $J - v$ such that $N(v) \subseteq N(u)$, and $H = J - v$. We say that the graph J *folds onto* H. Note that, for loop-free graphs, the definition implies that u and v are non-joined. We write $H \preccurlyeq J$ if there is a nonnegative integer m, graphs $H_0 = H, H_1, \ldots, H_m = J$, and nodes $v_0, \ldots, v_{m-1} \in V(J)$ so that $H_i \preccurlyeq_{v_i} H_{i+1}$ for all $0 \le i \le m-1$. Note that the relation $\preccurlyeq$ is an order relation on the class of all finite graphs. We name this the *folding order*. For example, $K_2 \preccurlyeq T$ where T is a tree, while two cliques of different orders are incomparable in the folding order. The relation $\preccurlyeq$ is a form of dismantling used to characterize certain vertex pursuit games. The graphs above K_2 in the folding order are sometimes called *dismantlable graphs*; see [4].

We extend the folding order to countable graphs as follows. Let H and J be countable graphs. The relation $H \preccurlyeq_v J$ is defined exactly as in the finite case. Fix I as either $\mathbb{N}$ or one of the sets $\{0, 1, \ldots n\}$, where $n \in \mathbb{N}$. We write $H \preccurlyeq J$ if there exists a sequence of countable graphs $(H_t : t \in I)$ so that $H_0 = H$, $H_t \preccurlyeq_v H_{t+1}$ for all $t \in I$, and $J = \lim_{t\to\infty} H_t$ if $I = \mathbb{N}$, or $J = H_n$ if I is of the form $\{0, 1, \ldots n\}$. For example, $K_2 \preccurlyeq G$, where G is the infinite one-way path. Note that for all $t > 0$, $H \preccurlyeq R_t$, and $R_t \preccurlyeq R_{t+1}$. Hence, $H \preccurlyeq R_H$.

Assume that H is connected and nontrivial. It is straightforward to see that the graph R_H consists of a unique infinite connected component, along with infinitely many isolated nodes. We use the notation C_H for this infinite component. We use the notation C_t for the unique nontrivial connected component of R_t in R_H. Hence, $C_H = \lim_{t\to\infty} C_t$, where $C_0 \cong H$, and $C_t \preccurlyeq_v C_{t+1}$ for all $t \in \mathbb{N}$. Note that each $v \in V(C_{t+1}) \backslash V(C_t)$ is joined to some node in C_t. The following theorem ties together the relation $\preccurlyeq$, the graph C_H, and the locally e.c. property.

Theorem 2 *Let H be a fixed finite nontrivial connected graph. If G is a countable connected locally e.c. graph such that $H \preccurlyeq G$, then $G \cong C_H$.*

Proof: Suppose, without loss of generality, that $G = \lim_{t\to\infty} G_t$, where $G_0 \cong H$, and $G_t \preccurlyeq_v G_{t+1}$ for all $t \in \mathbb{N}$. Since G is connected, each $v \in V(G_{t+1})\backslash V(G_t)$ is joined to some node in G_t.

We define an isomorphism $f : C_H \to G$ inductively as follows. Let $f_0 : C_0 \to G_0$ be any fixed isomorphism. As the induction hypothesis, suppose that for a fixed integer $t \geq 0$, there is a finite induced subgraph J_t of G containing G_t along with an isomorphism $f_t : C_t \to J_t$ extending f_0. (Note that we do not claim here that J_t is of the form G_s for some $s \geq 0$.)

Enumerate all pairs (y, S) where y is a node of J_t and $S \subseteq N_{J_t}(y)$ is nonempty as $\{(y_i, S_i) : 1 \leq i \leq k_t\}$. By the locally e.c. property for G, there is a node $x_{y_1,S_1} \in V(G)\backslash V(J_t)$ that is joined to S_1 and to no other nodes of J_t. A straightforward (and therefore, omitted) inductive argument supplies, for all $j \in \{1 \dots k_t\}$, a node x_{y_j,S_j} in

$$V(G)\backslash(V(J_t) \cup \{y_1, \dots, y_j\})$$

that is joined to S_j and to no other node of $V(J_t) \cup \{y_1, \dots, y_j\}$. Let

$$T = \{x_{y_i,S_i} : 1 \leq i \leq k_t\} \subseteq V(G).$$

Define J_{t+1} to be the subgraph of G induced by $V(J_t) \cup T$. Note that T is an independent set of nodes. By the definition of C_{t+1}, J_{t+1} is isomorphic to C_{t+1} by an isomorphism extending f_t. The unique node $v \in V(G_{t+1})\backslash V(G_t)$ is of the form $x_{y,S}$ for some y and S in G_t. As G_t is an induced subgraph of J_t, we have that $v \in V(J_{t+1})$. Hence, G_{t+1} is contained in J_{t+1}, and the induction step is complete.

Define $f : C_H \to G$ by $f = \bigcup_{t\in\mathbb{N}} f_t$. The map f is an embedding as each map f_t is an isomorphism, and it is onto by construction. Hence, f is an isomorphism. □

If $p \in (0, 1)$, then the classic results of Erdős, Rényi [8] imply that with probability 1, an infinite limit of $G(n, p)$ graphs is isomorphic to R. Perhaps surprisingly, for the duplication model there is a similar result, replacing R by C_H. The next corollary follows directly by Theorems 1 and 2.

Corollary 3 *Fix $p \in (0, 1)$ and H a finite nontrivial connected graph. With probability 1 a limit graph $G = \lim_{t\to\infty} G_t$ generated by the duplication model is isomorphic to the disjoint union of C_H and a set I of isolated nodes, where $|I|$ is countable.*

Following Brightwell and Winkler in [4], we say that a graph J is *stiff* if there is no proper induced subgraph H with the property that $H \preccurlyeq J$. By Theorem 4.4 of [4], each finite graph H contains a unique (up to isomorphism graph) stiff induced subgraph, written $c(H)$, so that $c(H) \preccurlyeq H$. We refer to $c(H)$ as the *stiff-core* of H. The stiff-core determine the isomorphism types of our limit graphs as follows.

Corollary 4 *Fix finite connected nontrivial graphs H and J. If $H \preccurlyeq J$, then $R_H \cong R_J$. In particular, $R_H \cong R_{c(H)}$.*

Our next result demonstrates how the graphs R_H play the role of "minimal" graphs for the locally e.c. property.

Theorem 5 *Fix H a finite graph, and let G be a locally e.c. graph. Then $R_H \leq G$ if and only if $H \leq G$.*

A vertex mapping $f : G \to H$ is a *homomorphism* if $xy \in E(G)$, implies that $f(x)f(y) \in E(H)$. We write $G \to H$ to denote that G admits a homomorphism to H without reference to a specific mapping. See the excellent book [9] for more background on graph homomorphisms. Folding gives rise to homomorphisms as follows.

Lemma 6 *Let G and H be countable graphs. If $H \preccurlyeq G$, then $G \to H$.*

The converse of Lemma 6 does not follow (consider H as K_2 and G as C_6). However, the following theorem establishes an interesting connection between graph homomorphisms and the induced subgraphs of R_H.

Theorem 7 *Fix n a positive integer, and let G and H be finite graphs. Then $G \leq R_H$ if and only if $G \to H$.*

Proof: For the forward direction, suppose that $G \leq R_H$. Then $G \leq R_t$ for some $t \geq 0$, and so $G \to R_t$. Since $H \preccurlyeq R_t$, by Lemma 6, we have that $R_t \to H$. Hence, $G \to H$ by the transitivity of the homomorphism relation.

For the converse, we introduce an auxiliary graph construction. Fix $f : G \to H$ a homomorphism. Assume $V(G)$ and $V(H)$ are disjoint, and define a graph $H(G, f)$ to have nodes $V(G) \cup V(H)$, and edges

$$E(G) \cup E(H) \cup \{xy : x \in V(G), y \in V(H), \text{and } f(x)y \in E(H)\}.$$

We refer to the induced copy of H in $H(G, f)$ as H'. We proceed by induction on $|V(G)|$ to show that $H(G, f) \leq R_H$. Since $G \leq H(G, f)$, the proof of the converse will follow. Note that, if $|V(G)| = 1$, then $H(G, f)$ is isomorphic to the disjoint union of H and K_1. The base case follows. The induction hypothesis is that if $|V(G)| = n$, where $n \geq 1$ is fixed, then $H(G, f)$ is an induced subgraph of R_H with H' the copy of H at $t = 0$.

Let $|V(G)| = n+1$, and fix $x \in V(G)$. By induction hypothesis, $H(G-x, f \restriction (G-x))$ is a subgraph S of R_H with H' the copy of H at $t = 0$. By the definition of $H(G, f)$, all the neighbours of x in $H(G, f)$ are also neighbours of the node $f(x)$ in H'. Note that $N(x) \subseteq V(S)$. By the locally e.c. property of R_H, there is a node z of R_H joined to the nodes of $N(x)$ in S and to no other nodes of S. Adding z to S in R_H will give an induced subgraph of R_H which is isomorphic to $H(G, f)$, while H' is unchanged. This completes the induction. □

Theorem 7 characterizes the *age* of R_H (that is, the class of isomorphism types of finite induced subgraphs of R_H). In addition, it supplies us with structural information about R_H.

Corollary 8
1. *For a fixed finite graph H, all countable H-colourable graphs embed in R_H.*
2. *The clique and chromatic numbers of R_H equal the clique and chromatic numbers of H, respectively.*

As one consequence, $R_{K_2} \leq R_{C_6}$ and $R_{C_6} \leq R_{R_2}$. Observe however, that $R_{K_2} \not\cong R_{C_6}$ as R_{C_6} does not fold onto K_2.

In the duplication model, the neighbourhood of a new node will always be a subset in the neighbourhood of an existing node. While this is sufficient for the modelling of protein-protein interaction networks, for other applications it is desirable to extend the model to allow for a number of random edges. Let $\rho : \mathbb{N} \to \mathbb{N}$ be a monotone increasing function. The duplication model can be generalized by adding a second step, where at each time t a set of $\rho(t)$ nodes is selected u.a.r., and edges from each node in this set to the new node are added. There is no unique limit in this case, but it can be shown that there is a unique minimal infinite graph which, with probability 1, is contained in any infinite limit. These minimal graphs are graphs that extend the idea behind the creation of R_H, and are called $R_H^{(n)}$. They are defined as follows.

Fix a finite graph H. Let $R_0 \cong H$. Assume that R_t is defined and is finite. For each node $y \in V(R_t)$, and each subset $X \subseteq N(y)$, add a new node $z_{y,X}$ joined only to X. For each subset of Y of nodes with cardinality at most n add a new node z_Y joined only to Y. This gives the graph R_{t+1} which contains R_t as an induced subgraph. Define $R_H^{(n)} = \lim_{t\to\infty} R_t$.

Results similar to those described above hold for $R_H^{(n)}$ for an appropriate adjacency property (namely, the n-e.c. property). These results will be described more fully in the journal version of this extended abstract.

References

[1] B. Bollobás, O. Riordan, Mathematical results on scale-free graphs, In: *Handbook of graphs and networks* (S. Bornholdt, H. Schuster, eds.) Wiley-VCH, Berlin (2002) 1-34.

[2] A. Bonato, A survey of models of the web graph, In: *Proceedings of Combinatorial and Algorithmic Aspects of Networking*, 2004.

[3] A. Bonato, J. Janssen, Infinite limits of copying models of the web graph, *Internet Mathematics* **1** (2004) 193-213.

[4] G. Brightwell, P. Winkler, Gibbs measures and dismantlable graphs, *J. Comb. Theory Ser. B* **78** (2000) 141-169.

[5] 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.

[6] P.J. Cameron, The random graph revisited, In: *European Congress of Mathematics* Vol. I (C. Casacuberta, R. M. Miró-Roig, J. Verdera and S. Xambó-Descamps, eds.), Birkhauser, Basel (2001) 267-274.

[7] F.R.K. Chung, G. Dewey, D.J. Galas, L. Lu, Duplication models for biological networks, *J. Comput. Biol.* **10** (2003) 677-688.

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

[9] P. Hell, J. Nešetřil, *Graphs and homomorphisms*, Oxford University Press, Oxford, 2004.

[10] J. Kleinberg, R. Kleinberg, Isomorphism and embedding problems for infinite limits of scale-free graphs, In: *Proceedings of ACM-SIAM Symposium on Discrete Algorithms*, 2005.