

The strong-connectivity of word-representable digraphs

Edward J. L. Bell

Department of Maths & Stats
Lancaster University, UK
e.bell@comp.lancs.ac.uk

Paul Rayson

School of Computing and Communications
Lancaster University, UK
p.rayson@lancs.ac.uk

Damon Berridge

Department of Maths & Stats
Lancaster University
d.berridge@lancs.ac.uk

January 11, 2013

Abstract

A word-graph G_ω is a digraph represented by a word ω such that the vertex-set $V(G_\omega)$ is the alphabet of ω and the edge-set $E(G_\omega)$ is determined by non-identical adjacent letter pairs in ω . In this paper we study the strong-connectivity of word-graphs. Our main result is that the number of strongly connected word-graphs represented by ℓ -words of over an n -alphabet can be expressed via the recurrence relation $T(\ell, n)$ on the Stirling numbers of the second kind using a link between word partitions and digraph connectivity.

$$T(\ell, n) = \binom{\ell-1}{n} + \sum_{j=0}^{\ell-2} \sum_{m=0}^{n-2} \binom{j}{m} T(\ell-j-1, n-m)(n-m-1)$$

1 Introduction

A digraph D is representable if D can be mapped to a word in some manner. In this paper we study a particular form of edge-set encoding for word-representable digraphs that has close links with discrete stochastic processes.

Definition 1. A word-graph G_ω is a simple digraph associated with an ℓ -word ω over an n -alphabet for $\ell > n > 0$ such that the vertex set of G_ω , $V(G_\omega)$, is the alphabet of ω and the directed edge-set of G_ω , $E(G_\omega)$, is encoded from ω using the set of non-identical adjacent letter pairs:

$$\begin{aligned} V(G_\omega) &= A(\omega) \\ E(G_\omega) &= E(\omega) = \{(\omega_1, \omega_2), (\omega_2, \omega_3), \dots, (\omega_{\ell-1}, \omega_\ell)\} \end{aligned}$$

such that for $\{\omega_i, \omega_{i+1}\} \in E(G_\omega)$ it holds that $\omega_i \neq \omega_{i+1}$

The edge-set encoding in Definition 1 is of particular note due to its close links with a range of stochastic word-based processes such as language, music, DNA and protein sequences. This form of edge-set encoding is also closely related to the concept of digraph path decomposition[1]. Other forms of edge-set encoding have been used in the literature for both word-representable graphs[4, 7] and word-representable digraphs[5]. These existing encodings are used for very specific applications (such as semigroup or DNA sequence analysis) whilst word-graphs have a wide range of general applications.

The focus of this paper is on determining the criteria required for a word to represent a strongly-connected digraph and how many words, out of all ℓ -words over an n -alphabet, represent strongly-connected digraphs. Strong-connectivity is a vital concept for word-graph theory because every strongly-connected word-graph, or component, can be represented by a single word[3].

Definition 2. A digraph D is strongly-connected if there exists a path from every $v_i \in V(D)$ to every $v_j \in V(D)$.

This paper is organised into four sections. The remainder of the first section details previous results and the notation used in this paper. The second section describes our results and the third section contains proofs of these results. Future work is detailed in the final section.

1.1 Previous results

We say a digraph D is representable, *i.e.* is a word-graph, if there exists a word ω that encodes D in the manner specified by Definition 1. Theorem 1 provides the criterion to be satisfied for a digraph to be representable

although there are other ways to characterise representability such as forbidden substructures or condensation forms[3].

Theorem 1. *A digraph D is a word-graph (representable) if there exists a directed path P from $v_i \in V(D)$ to $v_j \in V(D)$ such that P transverses each edge of D at least once. The sequence of vertices visited by P is a representational word of D .*

Corollary 1. Representable digraphs are weakly-connected.

Corollary 2. Strongly-connected digraphs are representable.

In the same manner that digraphs are characterised by vertex and edge set cardinality, word-graphs are characterised by word length and alphabet cardinality. The number of word-graphs represented by words of length ℓ over an alphabet of cardinality n can be expressed using Stirling numbers of the second kind[2].

Theorem 2. *The set of ℓ -words over an n -alphabet is of cardinality $n! \left\{ \begin{smallmatrix} \ell \\ n \end{smallmatrix} \right\}$.*

The term $\left\{ \begin{smallmatrix} \ell \\ n \end{smallmatrix} \right\}$ is a Stirling number of the second kind, or the number of ways of partitioning a set of cardinality ℓ into n subsets. If p is an n -partition of ℓ then the word corresponding to p would be: $\omega = p_1 p_2 \dots p_n$, for some ordering (*i.e.* lexicographical) of the partition members. The $n!$ term in Theorem 2 accounts for the permutation of the partitions to reflect the order in which they can occur in a word. We refer to the set of ℓ -words over an n -alphabet as a word-graph family $\mathfrak{W}(\ell, n)$. By Corollary 2, every strongly-connected digraph is represented by a word in some word-graph family.

Finally we define Menger's Theorem[6] which is used in the proofs in Section 3.

Theorem 3. *The edge-connectivity of a digraph is equal to the minimum number of edge-disjoint paths between any vertex pair.*

1.2 Notation and terminology

An ℓ -word ω over an n -alphabet is an ordered sequence of symbols of length $\ell = |\omega|$: $\omega = \omega_1 \dots \omega_\ell$ such that $\omega_i \in A(\omega)$ for $|A(\omega)| = n$ where $A(\omega)$ is the alphabet of ω . A circular word is one for which $\omega_1 = \omega_\ell$ and a trivial word is one for which $|A(\omega)| = 1$.

A factor of a word ω is any $\omega_j \dots \omega_k$ contiguous subword. A partition, or a factorisation, p of a word ω is a k -tuple for $1 \leq k \leq \ell$ of factors such that $\omega = p_1 p_2 \dots p_k$. A disjoint partition of ω is a partition such that, for each $p_i \neq p_j$ pair of partitions, the alphabets of p_i and p_j are disjoint: $A(p_i) \cap A(p_j) = \emptyset$. A minimal disjoint partition of ω is a partition of ω with the smallest cardinality such that the partitions remain disjoint.

The edge-connectivity of a word-graph is $\lambda(G_\omega)$ where $\lambda(G_\omega)$ is the cardinality of the edge-cut set $S \subset E(G_\omega)$ such that S is the smallest set of edges that disconnects G_ω when removed.

2 Results

The initial results in this paper take the form of three lemmas concerning the structure of words that represent strongly-connected digraphs. The three lemmas build upon one another. The final lemma is used to prove a theorem on the number of words that represent strongly-connected word-graphs in a word-graph family. The three lemmas are based upon the link between word-graph strong-connectivity and edge-connectivity.

Lemma 1. *Iff $\lambda(G_\omega) > 1$ then G_ω is strongly-connected.*

A bridge is an unidirectional edge in G_ω that when removed disconnects G_ω . When such a bridge exists in G_ω we say G_ω is 1-edge-connected or that $\lambda(G_\omega) = 1$. Lemma 1 proves that word-graphs with bridges are the only word-graph which are not strongly-connected. This edge-connectivity property holds due to the way in which edges are encoded from the representational word (Definition 1). Using Lemma 1 we prove Lemma 2 which characterises the structural properties of words that represent strongly-connected digraphs.

Lemma 2. *Iff $\lambda(G_\omega) > 1$ then ω cannot be partitioned into disjoint factors.*

See Subsection 1.2 for a definition of *disjoint factors*. Lemma 2 is restated by Lemma 3 in a manner that facilitates the enumeration of strongly-connected word-graphs.

Lemma 3. *The number of ℓ -words over an n -alphabet, with a unique factorisation, that represent word-graphs with $\lambda(G_\omega) > 1$ is the number of n -partitions of an ℓ -set such that no proper subset of parts has a union equal to the subset $\{1, 2, \dots, j\}$ for $j < \ell$.*

Via the three structural lemmas, the main enumerative result is proved in Theorem 4 using a recurrence relation on Stirling numbers of the second kind.

Theorem 4. *There exist $\varphi(\ell, n)$ strongly-connected word-graph members of the family $\mathfrak{W}(\ell, n)$ where:*

$$\begin{aligned}\varphi(\ell, n) &= n!T(\ell, n) \\ T(\ell, n) &= \left\{ \begin{matrix} \ell - 1 \\ n \end{matrix} \right\} + \sum_{j=0}^{\ell-2} \sum_{m=0}^{n-2} \left\{ \begin{matrix} j \\ m \end{matrix} \right\} T(\ell - j - 1, n - m)(n - m - 1)\end{aligned}$$

The recurrence in Theorem 4 was originally derived by Zabrocki[8] for an algebraic problem but, as yet, no formal proof has been published.

3 Proofs of results

In this section we use the previous results from Subsection 1.1 to prove the theorems and lemmas introduced in Section 2. We start with a characterisation of the structure of strongly-connected word-graphs using edge-connectivity.

Lemma 1. *Iff $\lambda(G_\omega) > 1$ then G_ω is strongly-connected.*

Proof. A strongly-connected digraph D , by Definition 2, has two edge-disjoint paths between every vertex pair thus, by Menger's Theorem[6], D is always 2-edge-connected. The remainder of this proof is dedicated to showing that if $\lambda(G_\omega) > 1$ then G_ω is strongly-connected.

Let K_n be the complete digraph of order n such that its vertices are labelled with the first n integers. Let ω be a walk on K_n and let G_ω be the word-graph represented by ω . We take bidirectional edges in G_ω to be pairs of unidirectional edges with opposing orientations. If the walk ω resides in vertex $v_j \in V(K_n)$ then the walk can take one of two actions:

- i. transition from v_j to v_{j+k} for $j + k \leq n$
- ii. transition from v_j to v_{j-k} for $j - k \geq 1$

Assume, without loss of generality, that ω starts on v_1 and, for the first $n - 1$ steps, uses action i so that G_ω is weakly-connected (Corollary 1) and, by Menger's Theorem[6], is 1-edge-connected (Figure 1).

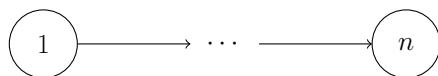


Figure 1: G_ω after $n - 1$ steps on K_n

To ensure G_ω is 2-edge-connected, the walk ω must continue. The walk cannot use action i because it already resides in v_n therefore it must use action ii to make the transition from v_n to v_{n-j} (Figure 2).

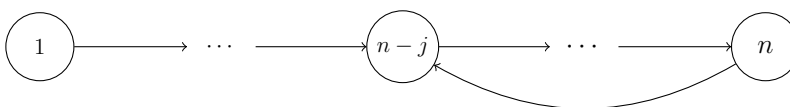


Figure 2: G_ω after n steps on K_n

G_ω now contains the strongly-connected component $C = \{v_n, \dots, v_{n-j}\}$. By Menger's Theorem[6], C is 2-edge-connected. To ensure the entirety of G_ω is 2-edge-connected, ω must again make the transition outside C so that C becomes larger still. If C does not encompass all nodes in K_n then G_ω cannot be 2-edge-connected because G_ω contains a bridge. Therefore the walk ω must continue in this manner until the strong-component C encompassed all vertices so that G_ω is 2-edge-connected and strongly-connected (Figure 3).

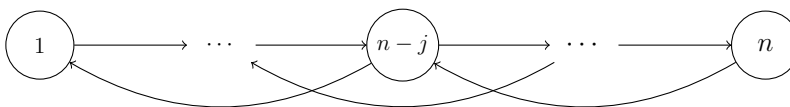


Figure 3: G_ω after $n + k$ steps on K_n

Thus we have shown that 2-edge-connectivity is a necessary and sufficient condition for the strong-connectivity of word-graphs. \square

Next we show that the structure of words that represent strongly-connected digraphs is characterised by the digraph's *bridgelessness* (2-edge-connectedness).

Lemma 2. *Iff $\lambda(G_\omega) > 1$ then ω cannot be partitioned into disjoint factors.*

Proof. Let p be a k -partition of the ℓ -word ω . If p is a disjoint partition then $2 \leq k \leq n$. If $k > n$ then, by the pigeon-hole principle, there must

be some pair of non-disjoint partitions because $|A(\omega)| = n$. If $k = 1$ then $p_1 = \omega$ and thus cannot be disjoint.

Assume p is a disjoint k -partition of ω for $2 \leq k \leq n$. For any p_i, p_{i+1} partition pair, contiguous in ω , of length u and v respectively, there exists an edge between the vertices $p_{i,u}$ and $p_{i+1,v}$ in G_ω . Given that all p_i partitions are disjoint, vertex $p_{i,u}$ occurs in no other partition but p_i and vertex $p_{i+1,v}$ occurs in no other partition but p_{i+1} . Therefore there exists only a single unidirectional edge between $p_{i,u}$ and $p_{i+1,v}$ (Figure 4).

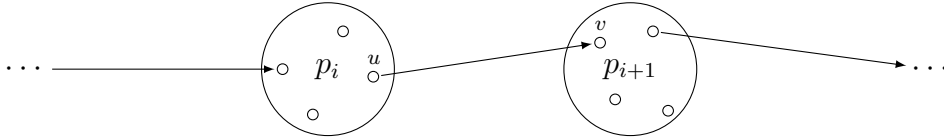


Figure 4: A word-graph represented by a contiguous disjoint partition

Furthermore, because every p_{i-j}, p_{i+1+m} partition pair is disjoint, there can be no path between $p_{i,u}$ and $p_{i+1,v}$ except the $\{p_{i,u}, p_{i+1,v}\}$ edge. If that edge were to be removed then G_ω would become disconnected. As such, if ω can be partitioned into $k > 1$ disjoint factors then G_ω contains a bridge and $\lambda(G_\omega) = 1$. If ω cannot be partitioned into disjoint factors then G_ω cannot contain a bridge and $\lambda(G_\omega) > 1$. \square

Corollary 3. If a word ω can be minimally partitioned into k disjoint factors then G_ω contains $k - 1$ bridges.

Lemmas 1 and 2 show that words that represent strongly-connected digraphs cannot be partitioned into disjoint factors. Lemma 3 restates these results in a manner that facilitates the enumeration of strongly-connected word-graphs.

In Theorem 2, the $n!$ term of $n! \binom{\ell}{n}$ accounts for factor permutation. In Lemma 3 the $n!$ term will be temporary ignored to make the proof simpler such that we only deal with the set of ℓ -words over an n -alphabet with a unique factorisation, *i.e.* we are not concerned with the order of those factors.

Lemma 3. *The number of ℓ -words over an n -alphabet, with a unique factorisation, that represent word-graphs with $\lambda(G_\omega) > 1$ is the number of n -partitions of an ℓ -set such that no proper subset of parts has a union equal to the subset $\{1, 2, \dots, j\}$ for $j < \ell$.*

Proof. Let p be an n -partition of an ℓ -set with members equal to the first ℓ non-zero integers. Let each p_i be mapped to an alphabet member $a_i \in A(\omega)$ of some word ω . Let the j -th integer member of p_i be the index of letter $a_i \in A(\omega)$ in ω . This construction gives an ℓ -word over an n -alphabet:

$$\omega_{p_{ij}} = h(p_i) = a_i \quad (3.1)$$

The number of n -partitions of an ℓ -set is $\binom{\ell}{n}$, which is the same as the number of ℓ -words over an n -alphabet with a unique factorisation.

If a proper $0 < k < \ell$ subset of p has a union equal to the subset $\{1, 2, \dots, j\}$ for $j < \ell$ then p is deemed reducible. A word constructed from a reducible partition is called partition-reducible. For a partition-reducible ℓ -word, the factor $\omega_1 \dots \omega_j$ is over the alphabet $\{a_1, \dots, a_k\}$ and the factor $\omega_{j+1} \dots \omega_\ell$ is over the alphabet $\{a_{k+1}, \dots, a_n\}$. Therefore there exists a disjoint partition of ω between ω_j and ω_{j+1} . Thus partition-reducible words cannot represent strongly-connected word-graphs because the word-graphs are 1-edge-connected (Lemmas 1 and 2).

If a partition is not reducible, it is irreducible. Let ω be a partition-irreducible word formed from the partition p as described by Equation 3.1. Assume $1 \in p_1$ and because p is irreducible, the union of p_1 with any proper subset of $p - \{p_1\}$ cannot be equal to $\{1, \dots, j\}$. Therefore there exists no disjoint partition of ω . Thus partition-irreducible words always represent strongly-connected word-graphs because the word-graphs are at least 2-edge-connected (Lemmas 1 and 2). \square

Via Lemma 3, a restatement of Lemma 2, we can count the number of strongly-connected word-graphs in $\mathfrak{W}(\ell, n)$.

Theorem 4. *There exist $\varphi(\ell, n)$ strongly-connected word-graph members of the family $\mathfrak{W}(\ell, n)$ where:*

$$\begin{aligned} \varphi(\ell, n) &= n!T(\ell, n) \\ T(\ell, n) &= \binom{\ell-1}{n} + \sum_{j=0}^{\ell-2} \sum_{m=0}^{n-2} \binom{j}{m} T(\ell-j-1, n-m)(n-m-1) \end{aligned}$$

Proof. There are two ways to construct an irreducible ℓ -set n -partition. First, let p be a partition of $\{1, 2, \dots, \ell-1\}$ into n subsets. Assume that $1 \in p_1$ and add ℓ to p_1 . If p were a reducible partition then the union of, at maximum, $n-1$ subsets of p would form the set $\{1, \dots, \ell\}$. Besides 1 and

ℓ there exist $\ell - 2$ members of the ℓ -set distributed amongst the subsets of p . No subset can be empty thus in order to form the union $\{1, \dots, \ell\}$, p_1 and all remaining $\ell - 2$ elements must be in the union. Such a partition cannot be reducible because the union is not formed from a proper subset of the partition. Thus partitions constructed in such a manner are irreducible. There exist:

$$\left\{ \begin{array}{c} \ell - 1 \\ n \end{array} \right\} \quad (3.2)$$

partitions of this type.

For the second type, let p be an irreducible partition of $\ell - j - 1$ into $n - m$ and let \hat{p} be a partition of j into m . Let $p\hat{p}$ be the union of the two partitions. The previous irreducible partition type counted all partitions with $\ell \in p_1$ such that $1 \in p_1$. Putting $\ell \in \hat{p}_i$ makes $p\hat{p}$ reducible because $\bigcup p_i = \{1, 2, \dots, \ell - j - 1\}$. The only remaining option is to put ℓ into $p_i \in p$ such that $1 \notin p_1$, which ensures $p\hat{p}$ is irreducible because:

$$\begin{aligned} \bigcup p_i &= \{1, 2, \dots, \ell - j - 1, \ell\} \\ \bigcup \hat{p}_i &= \left(\bigcup p_i \right)^c = \{\ell - j, \dots, \ell - 1\} \\ \bigcup p\hat{p}_i &= \{1, \dots, \ell\} \end{aligned} \quad (3.3)$$

Such a partition is irreducible because the only possible union equal to $\{1, \dots, j\}$ is not formed from a proper subset of $p\hat{p}$ (Equation 3.3).

For some fixed j and m , there exist $T(\ell - j - 1, n - m)$ irreducible $\{1, \dots, \ell - j - 1\}$ partitions into $n - m$; $\left\{ \begin{array}{c} j \\ m \end{array} \right\}$ partitions of $\{\ell - j, \dots, \ell - 1\}$ into m ; and $n - m - 1$ ways of placing element ℓ into $p_i \in p$ such that p remains irreducible:

$$\left\{ \begin{array}{c} j \\ m \end{array} \right\} T(\ell - j - 1, n - m)(n - m - 1) \quad (3.4)$$

To count all such partitions, equation 3.4 has to be summed over all $0 \leq m \leq n - 2$ because $n - m - 1 < 1$ for $m > n - 2$ and over all $0 \leq j \leq \ell - 2$ because $\ell - j - 1 < 1$ for $j > \ell - 2$.

These two types of irreducible partition constructions must be disjoint because, for the first type, ℓ and 1 are always in the same subset whilst, for the second type, ℓ and 1 are never in the same subset. Likewise, all

irreducible partitions are of one of these two types. If ℓ and 1 are in the same subset of p then p is irreducible and must be of the first type because Equation 3.2 counts all such partitions. If $1 \in p_1$ and $\ell \in p_i$ for $i > 1$ and p is irreducible then p is a partition of type two. If p were not of type two then ℓ must be in its own subset or in a subset with members of $\{j+1, \dots, \ell-1\}$ thus the union of a subset of $\{p_1, \dots, p_{i-1}\}$ would be equal to $\{1, \dots, j\}$ and so p would be reducible.

Combining Equations 3.2 and 3.4 gives the required recurrence relation.

$$T(\ell, n) = \binom{\ell-1}{n} + \sum_{j=0}^{\ell-2} \sum_{m=0}^{n-2} \binom{j}{m} T(\ell-j-1, n-m)(n-m-1)$$

The base cases for $T(\ell, n)$ are:

$$T(\ell, n) = 0 \text{ for } \ell \leq 0 \text{ and } \ell \leq n \quad (3.5)$$

$$T(\ell, 1) = 1 \quad (3.6)$$

The base cases in Equation 3.5 come from Definition 1. The base case in Equation 3.6 comes from the fact that the trivial ℓ -word ω with $|A(\omega)| = 1$ represents the singleton digraph which is strongly-connected. Including the $n!$ term, to account for factor permutation (see Theorem 2), gives the function $\varphi(\ell, n) = n!T(\ell, n)$ which is the number of strongly-connected word-graphs in $\mathfrak{W}(\ell, n)$. \square

4 Future work

We have shown how many members of a word-graph family, defined by a particular edge-set encoding with links to stochastic processes, are strongly-connected. Similar reasoning could be used for the number of word-graph with k strong-components.

Conjecture 1. *The number of ℓ -words over an n -alphabet, with a unique factorisation, that represent word-graphs with k strong-components is the number of n -partitions of an ℓ -set with k disjoint proper subsets of parts which have a union equal to the subset $\{j, j+1, \dots, j+m\}$. Each such union is an alphabet-disjoint partition in the corresponding representational word and is a strong-component in the corresponding word-graph.*

The word–graph family $\mathfrak{W}(\ell, n)$ includes isomorphic digraphs. The set of ℓ –words over an n –alphabet that represent non–isomorphic words–graphs can be thought of as a *fundamental* word–graph family. A standing question is: what is the cardinality of an arbitrary fundamental word–graph family and for what ℓ , when n is fixed, is the cardinality maximised? These topics are related to the concept of minimal representational word length[3].

Acknowledgements

The authors wish to thank Mike Zabrocki¹ of York University Canada for providing the recurrence relation in Theorem 4 and for giving a sketch of the proof, and thank the ESRC for funding this research.

References

- [1] B. Alspach and N. J. Pullman. Path decomposition of digraphs. *Bulletin of the Australian Mathematical Society.*, 10:421–427, 1974.
- [2] Edward J. L. Bell, Paul Rayson, and Damon Berridge. Word–graph enumeration with context–free grammars. In Review.
- [3] Edward J. L. Bell, Paul Rayson, and Damon Berridge. On the structural characterisation of word–graphs. In Review.
- [4] Sergey Kitaev and Artem Pyatkin. On representable graphs. *Journal of Automata, Languages and Combinatorics.*, 13(1):45–54, 2008.
- [5] Xianyue Li and Heping Zhang. Embedding on alphabet overlap digraphs. *Journal of Mathematical Chemistry.*, 47(1):62–71, 2010.
- [6] Karl Menger. Zur allgemeinen kurventheorie. *Fundamenta Mathematicae*, 10:95–115, 1927.
- [7] Arseny M Shur. On ternary square-free circular words. *The Electronic Journal of Combinatorics.*, 17(1), 2010.

¹zabrocki@mathstat.yorku.ca

- [8] N. J. A Sloane, 2003. Sequence in A087903 *The On-Line Encyclopedia of Integer Sequences* – <http://oeis.org/A087903>. Submitted by Mike Zabrocki.