Enumerating Simple Paths from Connected Induced Subgraphs

Pierre-Louis Giscard¹ · Paul Rochet²

Received: 25 January 2017 / Revised: 27 September 2018 / Published online: 25 October 2018
© Springer Japan KK, part of Springer Nature 2018

Abstract
We present an exact formula for enumerating the simple paths between any two vertices of a graph. Our formula involves the adjacency matrices of the connected induced subgraphs and remains valid on weighted and directed graphs. As a particular case, we obtain a relation linking the Hamiltonian paths and cycles of a graph to its dominating connected sets.

Keywords Directed graph · Self-avoiding walks · Simple cycles · Hamiltonian paths · Dominating sets · Inclusion–exclusion

1 Introduction

Counting simple paths, that is walks on a graph that do not visit any vertex more than once, is a problem of fundamental importance in enumerative combinatorics [12] with numerous applications, e.g. in sociology [7,14]. Several general purpose methods for counting simple paths and cycles have been discovered over the last 70 years, which make use of the inclusion–exclusion principle [2,3,5,6,10,11] or variants such as finite-difference sieves [4] and recursive expressions involving the adjacency matrix [1,9,13,14]. More rare but also worth mentioning are approaches using different tools such as zeon algebras [15] or immanantal equations [8].

In this paper, we derive a new formula for enumerating the simple paths from arbitrary vertices in a directed graph $G$. The result exploits the inclusion–exclusion principle over induced subgraphs, which led to the general purpose algorithm developed in [3]. We show that a simplification arises from factorizing induced subgraphs...
over their connected components, resulting in an expression that depends only on the connected induced subgraphs of $G$. A remarkable consequence is an expression that links the Hamiltonian paths of a graph to its dominating connected sets. A similar idea is used in [6] for the Travelling Salesman problem, where the reduction to dominating connected subsets is shown to be an improvement over the fastest known algorithms in graphs with bounded maximal degree.

### 2 Simple Path Enumeration

Let $G = (V, E)$ be a directed graph with vertex set $V = \{1, \ldots, n\}$ and edge set $E \subseteq V^2$, which may contain self-loops. The directed edge, or arc, from a vertex $i$ to a vertex $j$ is labeled $\omega_{ij}$. A path $p$ of length $k \geq 1$ is a sequence of $k$ contiguous arcs, that is, such that each new arc starts where the previous ended, e.g. $p = \omega_{ii_1}\omega_{i_1i_2}\cdots\omega_{i_{k-1}j}$. Paths appear naturally through analytical transformations of the labeled adjacency matrix $W$, with general term $W_{ij} = \omega_{ij}$ if $(i, j) \in E$ and $W_{ij} = 0$ otherwise. Precisely, paths of a given length $k \geq 1$ are enumerated in the $k$th power of $W$:

\[
(W^k)_{ij} = \sum_{\substack{p: i \to j \\
\ell(p) = k}} p, \quad i, j = 1, \ldots, n,
\]

where the sum runs over all paths $p$ of length $\ell(p) = k$ from $i$ to $j$ on $G$. Replacing $W$ by the (non-labeled) adjacency matrix $A$, $(A^k)_{ij}$ simply counts the number of paths of length $k$ from $i$ to $j$.

A path $p = \omega_{ii_1}\omega_{i_1i_2}\cdots\omega_{i_{k-1}j}$ is open if its end vertices $i, j$ are different and closed otherwise. A closed path is also called a cycle. An arc $\omega_{ij}$ is a path of length one from $i$ to $j$ while self-loops $\omega_{ii}$ and backtracks $\omega_{ij}\omega_{ji}$ are cycles of length one and two respectively.

A path $p = \omega_{ii_1}\omega_{i_1i_2}\cdots\omega_{i_{k-1}j}$ is simple if all indices $i, i_1, \ldots, i_{k-1}, j$ are different, with the possible exception $i = j$ if $p$ is a cycle. In the literature, variants of the inclusion–exclusion principle led to discovering exact formulas for counting simple paths and cycles on graphs. Exact formulas for small length paths [1,9,14] were later extended to paths of arbitrary length in e.g. [2,11], while a general expression is obtained in matrix form in [3].

Denote by $\pi_{ij}$ the enumeration of simple paths of any length from $i$ to $j$

\[
\pi_{ij} := \sum_{\substack{p: i \to j \\
p \text{ simple}}} p.
\]

For a matrix $M$ indexed by the vertices of the graph (typically, the adjacency matrix $A$ or the labeled version $W$), the restriction $M_S$ of $M$ to a non-empty subset $S$ of $V$ is
defined by
\[ M_{S,ij} = \begin{cases} M_{ij} & \text{if } i, j \in S, \\ 0 & \text{otherwise}, \end{cases} \]
i, j = 1, \ldots, n.

Substituting \( W \) for \( A \) in Eqs. (4) and (6) in [3], one obtains the following succinct formula for \( \pi_{ij} \),
\[
\pi_{ij} = \sum_{S \subseteq V, S \neq \emptyset} W_{S}^{\left| S \right| - \gamma_{ij}} (1 - W_{S})^{n - |S|} \]  
\[ ij \]
where \( \gamma_{ij} = 1 \) if \( i \neq j \) and 0 otherwise. Using the labeled adjacency matrix \( W \) enables to distinguish different paths. As a particular case, setting \( W = zA \) with \( z \) a formal variable creates the generating series of the number of simple paths by length. As noted in [3], the sum can of course be reduced to subsets \( S \subseteq V \) containing \( i, j \) since the term is zero otherwise.

Remark 1 This formula is meant for directed graphs. When working with undirected graphs with \( A \) (or \( W \)) symmetric, a simple path may be identified with its set of edges, not accounting for the direction. The corresponding sequence of edges and its reverse are thus seen as the same undirected object. In this context, undirected simple cycles of length \( k > 2 \) appear twice in the expression of \( \pi_{ii} \). For open paths however, the result remains valid as such with the particularity \( \pi_{ij} = \pi_{ji} \).

3 Enumerating Simple Paths from Weakly Connected Sets

A digraph is said to be weakly connected if replacing all its directed edges by undirected edges produces a connected undirected graph. The expression of \( \pi_{ij} \) can be reduced to a sum over weakly connected induced subgraphs of \( G \) owing to the simple property that the adjacency matrix of a disconnected digraph can be made block diagonal by an appropriate permutation of its rows and columns. Let \( G(S) \) denote the subgraph of \( G \) induced by \( S \subseteq V \). For all non-empty subsets \( S \) of \( V \), there is a unique partition \( C(S) = \{ C_1, \ldots, C_k \} \) dividing \( G(S) \) into non-empty weakly connected components such that \( G(S) = G(C_1) \cup \cdots \cup G(C_k) \). This partition verifies for all \( m \geq 1 \),
\[
W_{S}^{m} = W_{C_1}^{m} + \cdots + W_{C_k}^{m}. \]  
\( (3) \)

Let \( \mathcal{C} \) denote the set of non-empty subsets \( C \subseteq V \) for which the induced subgraph \( G(C) \) is weakly connected. For \( C \in \mathcal{C} \), the weak neighborhood \( N(C) \) of \( C \) in \( G \) is defined by
\[ N(C) = \{ i \in V \setminus C : \exists j \in C, \{(i, j), (j, i)\} \cap E \neq \emptyset \}. \]
This definition recovers the classical definition of neighborhood in undirected graphs.

**Theorem 1** For all \( i, j = 1, \ldots, n \),

\[
\pi_{ij} = \left[ \sum_{C \in \mathcal{C}} W_C^{N(C)} (I - W_C)^{|N(C)|} \right]_{ij},
\]

where we recall \( \gamma_{ij} = 1 \) if \( i \neq j \) and \( 0 \) otherwise.

**Proof** We show the proof for \( i = j \) (i.e. \( \gamma_{ij} = 0 \)) but the calculations are identical for \( i \neq j \). By (3), we have

\[
\sum_{S \neq \emptyset} W_S^{|S|} (I - W_S)^{n - |S|} = \sum_{S \neq \emptyset} \sum_{C \in \mathcal{C}(S)} W_C^{|S|} (I - W_C)^{n - |S|}
\]

\[= \sum_{C \in \mathcal{C}} \sum_{S : C \in \mathcal{C}(S)} W_C^{|S|} (I - W_C)^{n - |S|}.
\]

Fix \( C \in \mathcal{C} \). A set \( S \in \mathcal{S} \) such that \( G(C) \) is a weakly connected component of \( G(S) \) writes as \( S = C \cup T \) for \( T \subseteq V \setminus (C \cup N(C)) \). Thus,

\[
\sum_{S : C \in \mathcal{C}(S)} W_C^{|S|} (I - W_C)^{n - |S|} = \sum_{T \subseteq V \setminus (C \cup N(C))} W_C^{|C \cup T|} (I - W_C)^{n - |C \cup T|}.
\]

Since \( C \) and \( T \) are disjoint \( |C \cup T| = |C| + |T| \) and the above expression can be rewritten as

\[
W_C^{|C|} (I - W_C)^{|N(C)|} \sum_{T \subseteq V \setminus (C \cup N(C))} W_C^{|T|} (I - W_C)^{n - |C| - |N(C)| - |T|}.
\]

Let \( k = n - |C| - |N(C)| \), we obtain by regrouping the subsets \( T \) by size

\[
\sum_{T \subseteq V \setminus (C \cup N(C))} W_C^{|T|} (I - W_C)^{k - |T|} = \sum_{j=0}^{k} \binom{k}{j} W_C^{j} (I - W_C)^{k - j} = 1.
\]

Finally,

\[
\sum_{S \neq \emptyset} W_S^{|S|} (I - W_S)^{n - |S|} = \sum_{C \in \mathcal{C}} W_C^{|C|} (I - W_C)^{|N(C)|}
\]

and the result follows from Eq. (2). \( \square \)

The consequence for enumerating Hamiltonian paths follows by isolating the terms with exponent \( n - \gamma_{ij} \) in the expanded form of \( \pi_{ij} \). These terms correspond to the sets \( C \) such that \( |N(C)| = n - |C| \), i.e. dominating sets.
Corollary 1  Let $h_{ij}(G)$ denote the enumeration of Hamiltonian paths from $i$ to $j$ in $G$ and $\mathcal{D}$ the set of weakly connected dominating sets in $G$. For all $i, j = 1, \ldots, n$,

$$h_{ij}(G) = \left[ \sum_{D \in \mathcal{D}} (-1)^{n-|D|} W_D^{n-\gamma_{ij}} \right]_{ij}.$$ 

Remark 2  In undirected graphs, a factor $1/2$ must be applied to $h_{ii}(G)$ to enumerate the (undirected) Hamiltonian cycles in a graph of order $n > 2$.

Since a simple path of arbitrary length in the graph $G$ can be regarded as an Hamiltonian path in one of its connected induced subgraphs, the expression of $\pi_{ij}$ in Theorem 1 can be recovered from summing the $h_{ij}(G(C))$ over all connected induced subgraphs $G(C)$. Let $\mathcal{D}_C$ be the set of weakly connected dominating sets in $G(C)$, a connected subset $D$ belongs to $\mathcal{D}_C$ if and only if $D \subseteq C$ and $T := C \setminus D \subseteq N(D)$. Hence,

$$\sum_{C \in \mathcal{C}} h_{ij}(G(C)) = \sum_{C \in \mathcal{C}} \left[ \sum_{D \in \mathcal{D}_C} (-1)^{|C| - |D|} W_D^{|C| - \gamma_{ij}} \right]_{ij} = \sum_{D \in \mathcal{D}} \left[ \sum_{T \subseteq N(D)} (-1)^{|T|} W_D^{|D| + |T| - \gamma_{ij}} \right]_{ij} = \left[ \sum_{D \in \mathcal{C}} W_D^{D - \gamma_{ij}} \sum_{T \subseteq N(D)} (-W_D)^{|T|} \right]_{ij}.$$ 

Regrouping the subsets $T$ by size, we get

$$\sum_{T \subseteq N(D)} (-W_D)^{|T|} = \sum_{k=0}^{|N(D)|} \binom{|N(D)|}{k} (-W_D)^k = (1 - W_D)^{|N(D)|},$$

recovering $\sum_{C \in \mathcal{C}} h_{ij}(G(C)) = \pi_{ij}$ from Theorem 1 as expected.

The Hamiltonian cycles of $G$ appear on the diagonal values $h_{ii}(G)$. Since all are equal (and therefore equal to their mean value) the number of Hamiltonian cycles in $G$ is given by

$$\frac{1}{n} \sum_{D \in \mathcal{D}} (-1)^{n-|D|} \text{Tr}(A_D^n).$$

The reduction of the Travelling Salesman problem to dominating weakly connected sets has been investigated in [6, Theorem 3], where it proved to be a computational improvement for bounded degree graphs.

Acknowledgements  P.-L. Giscard is grateful for the financial support from the Royal Commission for the Exhibition of 1851. The authors are grateful to an anonymous referee for its many constructive remarks that helped improve the paper.
References

1. Alon, N., Yuster, R., Zwick, U.: Finding and counting given length cycles. Algorithmica 17, 209–223 (1997)
2. Bax, E.T.: Inclusion and exclusion algorithm for the hamiltonian path problem. Inf. Process. Lett. 47(4), 203–207 (1993)
3. Bax, E.T.: Algorithms to count paths and cycles. Inf. Process. Lett. 52(5), 249–252 (1994)
4. Bax, E., Franklin, J.: A finite-difference sieve to count paths and cycles by length. Inf. Process. Lett. 60(4), 171–176 (1996)
5. Björklund, A., Husfeldt, T., Kaski, P., Koivisto, M. (2009) Counting Paths and Packings in Halves. In: Fiat, A., Sanders, P. (eds) Algorithms - ESA 2009. ESA 2009. Lecture Notes in Computer Science, vol 5757, pp 578–586. Springer, Berlin, Heidelberg (2009)
6. Björklund, A., Husfeldt, T., Kaski, P., Koivisto, M.: The travelling salesman problem in bounded degree graphs. ACM Trans Algorithms 8(2) (2012)
7. Cartwright, D., Harary, F.: Structural balance: a generalization of heider’s theory. Psychol. Rev. 63, 277–293 (1956)
8. Cash, G.G.: The number of n-cycles in a graph. Appl. Math. Comput. 184(2), 1080–1083 (2007)
9. Harary, F., Manvel, B.: On the number of cycles in a graph. Matematický časopis 21, 55–63 (1971)
10. Karp, R.M.: Dynamic programming meets the principle of inclusion and exclusion. Oper. Res. Lett. 1(2), 49–51 (1982)
11. Khomenko, N.P., Golovko, L.D.: Identifying certain types of parts of a graph and computing their number. Ukrainskii Matematicheskii Zhurnal 24, 385–396 (1972)
12. Madras, N., Slade, G.: The self-avoiding walk. Springer Science & Business Media (2013)
13. Perepechko, S.N., Voropaev, A.N.: The number of fixed length cycles in an undirected graph. explicit formulae in case of small lengths. In: Mathematical Modeling and Computational Physics (MMCP2009), pp. 148–149 (2009)
14. Ross, I.C., Harary, F.: On the determination of redundancies in sociometric chains. Psychometrika 17, 195–208 (1952)
15. Schott, R., Staples, G.S.: Complexity of counting cycles using zeons. Comput. Math. Appl. 62(4), 1828–1837 (2011)