Notes on walk-counting

Gregory Quenell

September 1997

1 Walks and graphs

Let Γ be a finite, undirected, simple graph with n vertices. We will use the notation $i \sim j$ to indicate that two vertices i and j in Γ are joined by an edge. Let A be an adjacency matrix for Γ. That is, A is an $n \times n$ matrix with rows and columns indexed by the vertices of Γ, and

$$[A]_{ij} = \begin{cases} 1 & \text{if } i \sim j \\ 0 & \text{if } i \not\sim j. \end{cases}$$

We use the notation $[A]_{ij}$ to denote the entry in the i^{th} row and j^{th} column of the matrix A.

If m is a non-negative integer and i and j are vertices in Γ, then an m-walk from i to j is a sequence

$$v_0, e_1, v_1, e_2, v_2, \ldots, v_{m-1}, e_m, v_m$$

of vertices v_α and edges e_β such that

1. $v_0 = i$
2. e_α is incident on v_α and $v_{\alpha-1}$ for $1 \leq \alpha \leq m$
3. $v_m = j$.

The following is well-known ([1], Lemma 2.5, for example).

Theorem 1 *Let A be an adjacency matrix for a simple, undirected, finite graph Γ. For each non-negative integer m and each pair of vertices i and j in Γ, the number of m-walks from i to j is equal to $[A^m]_{ij}$.*

Proof We adopt the convention that a zero-walk can go only from a vertex to itself, and for each vertex there is exactly one such walk. Thus the number of zero-walks from i to j is 1 if $i = j$ and 0 if $i \neq j$. Since $A^0 = I$, the statement holds for $m = 0$.

There exists exactly one 1-walk from i to j if and only if $i \sim j$, and there are zero 1-walks from i to j otherwise. Since the entry $[A]_{ij}$ is 1 when $i \sim j$ and 0 when $i \not\sim j$, the statement holds for $m = 1$.

Now suppose $m \geq 2$ and the number of $(m-1)$-walks from i to any vertex k is equal to $[A^{m-1}]_{ik}$. Every m-walk from vertex i to vertex j may be viewed as an $(m-1)$ walk to a neighbor k of j, followed by a single step from k to j. In fact, the set of $(m-1)$-walks from i to neighbors of j is in one-to-one correspondence with the set of m-walks from i to j. Thus

$$\#(m\text{-walks from } i \text{ to } j) \quad = \quad \sum_{k \sim j} [A^{m-1}]_{ik}$$

where the sum is taken over all neighbors k of j. The expression on the right can be obtained by matrix multiplication; in fact

$$\sum_{k \sim j} [A^{m-1}]_{ik} \quad = \quad [A^{m-1}A]_{ij},$$

so that we get

$$\#(m\text{-walks from } i \text{ to } j) \quad = \quad [A^m]_{ij}$$

as required. ■

2 Counting geodesics

An m-walk will be called an *m-geodesic* if it does not traverse the same edge twice in succession. In order to count m-geodesics, we introduce the $n \times n$

matrices G_m. Let $[G_m]_{ij}$ be the number of m-geodesics from vertex i to vertex j.

Clearly, G_1 is the same as the adjacency matrix A. We will adopt the convention that G_0 is the identity matrix I. The matrix G_2, however, differs from A^2, since each diagonal entry $[A^2]_{ii}$ includes non-geodesic 2-walks from vertex i to each of its neighbors and back. That is, we have $A^2 = G_2 + D$, where D is a diagonal matrix with $[D]_{ii}$ equal to the degree of vertex i.

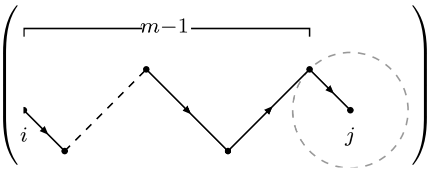
For $m \geq 3$, each m-geodesic from vertex i to vertex j can be viewed an an $(m-1)$-geodesic from i to a neighbor k of j, followed by a single step from k to j. Thus, all the m-geodesics from i to j are counted by the expression

$$\sum_{k \sim j} [G_{m-1}]_{ik} \tag{1}$$

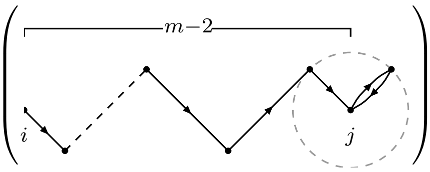
in which the sum is taken over all neighbors k of j.

Unfortunately, expression (1) also counts m-walks from i to j in which the antepenultimate vertex is j and the last step is a backtrack from a neighbor of j to j. We can represent the two different kinds of m-walks counted by expressing (1) with a picture equation

$$\sum_{k \sim j} [G_{m-1}]_{ik} = \#$$



$$+ \ \#$$



(2)

in which each shaded circle represents the set of neighbors k of vertex j.

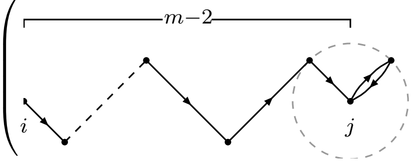
We have overcounted by including the m-walks in the second term on the right side of (2), each of which has a backtrack in its last two edges. We correct for this overcounting by subtracting the term

$$\sum_{k\sim j}[G_{m-2}]_{ij}, \tag{3}$$

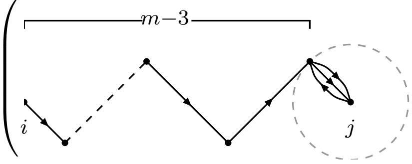
(note that the matrix index is ij, rather than ik; this is intentional) which accounts for all the m-walks from i to j which consist of an $(m-2)$-geodesic from i to j followed by a 2-walk from j to j, via some neighbor k of j.

Unfortunately, in addition to the m-walks we just described, expression (3) counts m-walks from i to j which contain an $(m-3)$-geodesic from i to a neighbor k of j followed by three traversals of the edge from j to k. Pictorially, we have

$$\sum_{k\sim j}[G_{m-2}]_{ij} = \#\Bigg($$



$$\Bigg) + \#\Bigg($$



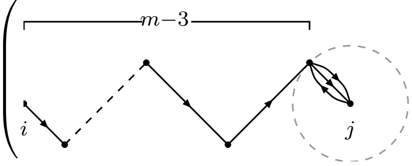
$$\Bigg). \tag{4}$$

By subtracting $\sum[G_{m-2}]_{ij}$ from $\sum[G_{m-1}]_{ik}$, we solve the problem of overcounting the walks in the second term on the right side of (2), but we have now oversubtracted the walks in the second term on the right side of (4). Since these walks correspond to $(m-3)$-geodesics from i to neighbors of j, we compensate for our oversubtraction by adding the expression

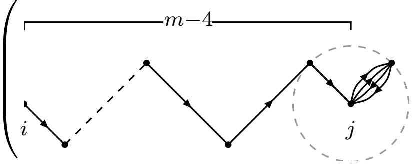
$$\sum_{k\sim j}[G_{m-3}]_{ik}. \tag{5}$$

Again, we have overcompensated, because expression (5), in addition to the walks we want, includes walks which pass through j at the $(m-4)^{\text{th}}$ step and then traverse an edge between j and some neighbor k four times in succession. That is,

$$\sum_{k\sim j}[G_{m-3}]_{ik} \quad = \quad \#\Bigg($$



$$\Bigg) + \#\Bigg($$



$$\Bigg). \qquad (6)$$

The pattern which emerges from this argument says that

$$[G_m]_{ij} \quad = \quad \sum_{k\sim j}\left([G_{m-1}]_{ik} - [G_{m-2}]_{ij} + [G_{m-3}]_{ik} - [G_{m-4}]_{ij} + \cdots\right) \qquad (7)$$

with $G_1 = A$ and $G_0 = I$.

It is not hard to check that the summations in (7) correspond to right-multiplication by certain matrices. Specifically, letting A denote the adjacency matrix and D denote the diagonal matrix whose ii^{th} entry is equal to the degree of vertex i, we have

Theorem 2 *The geodesic-counting matrices G_m for a graph with adjacency matrix A and degree matrix D satisfy the relations $G_0 = I$, $G_1 = A$, and for $m \geq 2$, $G_m = G_{m-1}A - G_{m-2}D + G_{m-3}A - G_{m-4}D + \cdots$.*

From Theorem 2, it follows immediately that for $m \geq 3$, we have

$$G_m - G_{m-2} \quad = \quad G_{m-1}A - G_{m-2}D. \qquad (8)$$

Now let $F(u)$ be an ordinary generating function for the matrices G_m. That is, let

$$F(u) \quad = \quad \sum_{m=0}^{\infty} G_m u^m.$$

Using relation (8) and the facts that $G_0 = I$, $G_1 = A$ and $G_2 = A^2 - D$, we determine that

$$F(u) - F(u)Au + F(D-I)u^2 \quad = \quad I - Iu^2.$$

We can solve this for $F(u)$, and thus in some sense determine all the matrices G_m at once. We have (see [2], Lemma 1)

Theorem 3 *The generating function $F(u)$ satisfies*

$$F(u) = (1-u^2)(I - Au + (D-I)u^2)^{-1}.$$

3 Graphs with edge lengths

Next we introduce a length function on the edges of our graph Γ. If i and j are vertices in Γ with $i \sim j$, then $\ell(i,j)$ will be some positive real number, which we will think of as the length of the edge connecting i and j. If $i \not\sim j$, then $\ell(i,j)$ is undefined.

We construct a length-adjacency matrix $\mathcal{A}$ for such a graph by introducing a dummy variable t and setting

$$[\mathcal{A}]_{ij} \quad = \quad \begin{cases} t^{\ell(i,j)} & \text{if} \quad i \sim j \\ 0 & \text{if} \quad i \not\sim j. \end{cases}$$

Let $v_0, e_1, v_1, e_2, \ldots, v_{m-1}, e_m, v_m$ be an m-walk from v_0 to v_m. We will say that the *length* of this m-walk is the positive real number

$$\ell(v_0, v_1) + \ell(v_1, v_2) + \cdots + \ell(v_{m-1}, v_m).$$

We have the following extension of Theorem 1.

Theorem 4 *For each non-negative integer m and each pair of vertices i and j, the entry $[\mathcal{A}^m]_{ij}$ is an expression of the form*

$$\sum_{r \in L} \alpha_r t^r$$

where L is the set of real numbers r for which there exists an m-walk of length r from i to j, and for each $r \in L$, α_r is equal to the number of m-walks of length r from i to j.

Proof We adopt the convention that all 0-walks have length 0. As such, for vertices i and j, we have exactly one 0-walk of length 0 if $i = j$ and no 0-walks otherwise. Since $\mathcal{A}^0 = I$, the statement holds for $m = 0$.

If $i \sim j$, then there is a 1-walk of length $\ell(i,j)$ from i to j. If $i \not\sim j$ then there are no 1-walks from i to j. By the definition of $\mathcal{A}$, the statement holds for $m = 1$.

Now suppose $m \geq 2$ and each entry $[\mathcal{A}^{m-1}]_{ij}$ has the form given in the statement of the theorem. An m-walk of length ℓ from i to j may be viewed as an $(m-1)$-walk to a neighbor k of j (which must have length $\ell - \ell(j,k)$) followed by a traversal of the edge from k to j. Since every m-walk from i to j must have a neighbor of j as its penultimate vertex, we can count them by summing over neighbors k of j. We get

$$\begin{aligned}
\#\begin{pmatrix} m\text{-walks from } i \text{ to } j \\ \text{of length } \ell \end{pmatrix} &= \sum_{k\sim j} \#\begin{pmatrix} (m-1)\text{-walks from } i \text{ to } k \\ \text{of length } \ell - \ell(j,k) \end{pmatrix} \\
&= \sum_{k\sim j} \text{coefficient of } t^{\ell-\ell(j,k)} \text{ in } [\mathcal{A}^{m-1}]_{ik} \\
&= \text{coefficient of } t^{\ell} \text{ in } \sum_{k\sim j} [\mathcal{A}^{m-1}]_{ik} t^{\ell(j,k)} \\
&= \text{coefficient of } t^{\ell} \text{ in } [\mathcal{A}^{m-1}\mathcal{A}]_{ij}
\end{aligned}$$

as required. ■

4 Geodesics with edge lengths

Next we wish to count geodesics of a given length on a graph equipped with an edge-length function. As in Section 2, we construct a sequence of matrices $\mathcal{G}_m$ whose entries will count m-geodesics. Specifically, the matrix entry $[\mathcal{G}_m]_{ij}$ will be an expression of the form

$$\sum_{r\in L} \alpha_r t^r$$

where L is the set of real numbers r for which there exists an m-geodesic of length r from i to j, and for each $r \in L$, α_r is the number of m-geodesics from i to j with length r.

Since each entry in $\mathcal{G}_m$ is a function of t, we will write $\mathcal{G}_m(t)$. We will also find it convenient in this section to treat the matrix $\mathcal{A}$ as a function of t, writing $\mathcal{A}(t)$.

The analogue of the degree matrix D from Section 2 is a diagonal matrix $\mathcal{D}(t)$ with

$$[\mathcal{D}(t)]_{jj} = \sum_{k\sim j} t^{\ell(j,k)}.$$

As in Section 2, our conventions about 0-walks make it easy to see that $\mathcal{G}_0(t)$ is the identity matrix, and by the definitions of $\mathcal{A}(t)$ and $\mathcal{G}_1(t)$, it is clear that $\mathcal{G}_1(t) = \mathcal{A}(t)$.

The relation between $\mathcal{G}_2(t)$ and $\mathcal{A}^2(t)$ is also familiar. If $i \neq j$, then $[\mathcal{G}_2(t)]_{ij} = [\mathcal{A}^2(t)]_{ij}$. Along the diagonal, however, $\mathcal{A}^2(t)$ includes terms which do not correspond to geodesics. Specifically,

$$[\mathcal{A}^2(t)]_{jj} = \sum_{k\sim j} t^{2\ell(j,k)}$$

with the sum taken over all neighbors k of j. Thus we get the relation

$$\mathcal{G}_2(t) = \mathcal{A}^2(t) - \mathcal{D}(t^2).$$

For $m \geq 3$, we compute $[\mathcal{G}_m(t)]_{ij}$ by considering $(m-1)$-geodesics to neighbors k of j. The expression

$$\sum_{k\sim j} [\mathcal{G}_{m-1}(t)]_{ik} t^{\ell(j,k)} \tag{9}$$

contains a term t^ℓ for every m-geodesic from i to j of length ℓ. Unfortunately, as before, this expression overcounts. It includes a term t^ℓ for each m-walk from i to j of length ℓ made up of an $(m-2)$-geodesic from i to j followed by a step out to a neighbor k of j and then a step back to k (cf. equation (2)).

To correct for this overcounting, we subtract the expression

$$\sum_{k\sim j}[\mathcal{G}_{m-2}(t)]_{ij}t^{2\ell(j,k)} \tag{10}$$

which includes a term t^ℓ for every m-walk of the type overcounted in (9). However, expression (10) does more than we wanted; its count also takes in m-walks from i to j in which the last three steps all traverse the same edge (cf. equation (4)). To compensate for this overcounting (or oversubtraction, really), we add the expression

$$\sum_{k\sim j}[\mathcal{G}_{m-3}(t)]_{ik}t^{3\ell(j,k)} \tag{11}$$

which includes a term t^ℓ for each m-walk from i to j of length ℓ which consists of an m-geodesic from i to a neighbor of j followed by three traversals of the same edge. Once again, though (cf. equation (6)), this expression carries along with it an undesirable side-effect, and we must continue subtracting and adding appropriate terms in order to compensate.

The pattern which emerges from this process is as follows. For $m \geq 3$, we have

$$\begin{aligned}[\mathcal{G}_m(t)]_{ij} &= \sum_{k\sim j}\Big([\mathcal{G}_{m-1}(t)]_{ik}t^{\ell(j,k)} - [\mathcal{G}_{m-2}(t)]_{ij}t^{2\ell(j,k)} \\ &\qquad + [\mathcal{G}_{m-3}(t)]_{ik}t^{3\ell(j,k)} - [\mathcal{G}_{m-4}(t)]_{ij}t^{4\ell(j,k)} + \cdots\Big)\end{aligned} \tag{12}$$

It is not hard to show that the summation in the odd terms corresponds to right multiplication by the matrix $\mathcal{A}$ evaluated at a suitable power of t, and the summation in the even terms corresponds to right multiplication by the matrix $\mathcal{D}$, also evaluated at a suitable power of t. Specifically, we have

Theorem 5 *Let $\mathcal{A}(t)$ and $\mathcal{D}(t)$ be the length-adjacency and length-degree matrices, respectively, of a graph Γ. The matrices $\mathcal{G}_m(t)$ which count the number of m-geodesics of any given length in Γ satisfy $\mathcal{G}_0(t) = I$, $\mathcal{G}_1(t) = \mathcal{A}(t)$, and for $m \geq 2$,*

$$\begin{aligned}\mathcal{G}_m(t) &= \mathcal{G}_{m-1}(t)\mathcal{A}(t) - \mathcal{G}_{m-2}(t)\mathcal{D}(t^2) \\ &\qquad + \mathcal{G}_{m-3}(t)\mathcal{A}(t^3) - \mathcal{G}_{m-4}(t)\mathcal{D}(t^4) + \cdots.\end{aligned}$$

References

[1] Norman Biggs. *Algebraic Graph Theory.* Cambridge University Press, Cambridge, UK, 1993.

[2] Audrey Terras and Harold Stark. Zeta functions of finite graphs and coverings. *Advances in Mathematics*, 121:124–165, 1996.