Math 336: Accelerated Advanced Honors Calculus
James Morrow
Spencer Peters
5/20/16

Review of *Random Walks on Graphs: A Survey* [1] with Applications to Thermodynamics And a Digression into the Nature of Harmonicity

Introduction

Consider a system with a set of states and a set of allowable transitions between pairs of states. We can represent this system as a graph, whose nodes are the states, and whose edges are possible transitions. In this model, the evolution of the system corresponds to a random walk on the graph.

A random walk on a graph is a sequence of nodes generated by starting at one node, and then moving to one of its neighbors selected at random, and then to one of its neighbors, and so on. Think shuffling a deck of cards. The current permutation of a deck of cards being shuffled evolves according to a random walk on a graph whose nodes are permutations, and whose edges relate permutations that can be obtained from each other via a single shuffling motion.

In this paper I'll summarize some of the results in Lovasz's 1993 paper *Random Walks on Graphs: A Survey* [1] that I found most interesting. First I'll derive the unique ordering of access times between nodes--some nodes are "easier to reach and harder to get away from", while others are "harder to reach and easier to get away from", and describe some consequences and interesting examples.

Next, I'll illustrate the deep connection between random walks and linear algebra, via the matrix of transition probabilities. Using eigenvalues and the spectral decomposition, I'll show that a random walk starting at any probability distribution tends to a unique stationary distribution.

Finally, I'll describe how random walks on graphs bear on statistical mechanics--in particular the fundamental assumption that all accessible states of a system are equally likely. I'll use my knowledge of the stationary distribution to prove the fundamental assumption for strongly reversible physical systems.

Let's dive in!

Basic Definitions

Let $G = (V, E)$ be a connected graph; let $|V| = n$ and $|E| =$ m. A random walk on G starts at a node $v0$; if at the t-th step we are at node vt, we move to a neighbor vi of vt with probability $Pt->i$. In what follows, (until I explicitly break this assumption in my discussion of statistical mechanics), we assume the graph is undirected (transitions can occur in either direction) and the probability $Pj->k$ of transitioning from node j is the same for each neighbor k; that is, $Pj->k$ = 1/out-degree(j).

The node $v0$ can be fixed, or it can be a random variable drawn from some initial distribution $D0$. We denote by Dt the distribution of the random walk after t steps.

The neat thing about this setup is that we can write down a matrix of transition probabilities M that acts on the vector Dt to produce a vector $Dt+1$. That is,

$Dt+1j = M\ Dt$ = Sum(1 <= i <= n of $Mij\ Dtj$)

where

Definition (Matrix of Transition Probabilities M): $Mij = Pi->j$

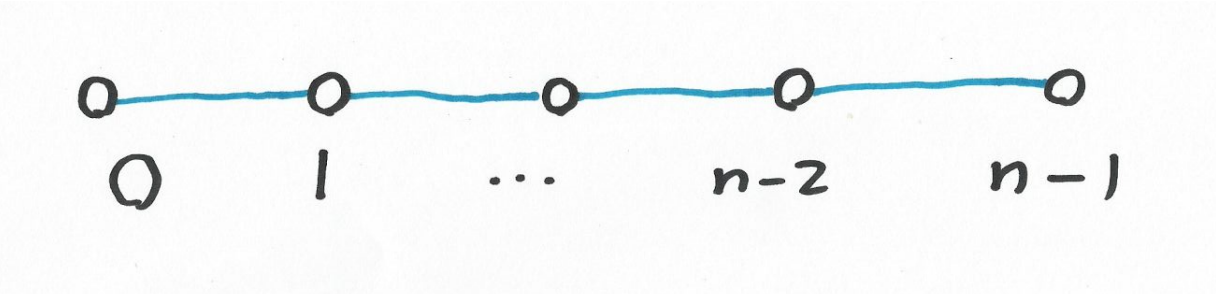
We'll return to this later.

Another very fundamental idea is that of access time.

Definition (Access Time): The access time $H(i, j)$between two nodes i and j is the expected number of steps before node j is visited starting from node i.

One can also look at the return time $R(k)$, which is the expected number of steps for a random walk starting at node k to return to k--this quantity will be important later. Or the commute time $K(i, j) = H(i, j) + H(j, i)$, which is the expected number of steps for a round-trip; or the cover time C-$D0$, which is the expected number of steps to reach every node starting from distribution $D0$. But we'll stick with the access time for now.

To illustrate the basic idea of access time, we'll compute the access time for two points on a **path** on nodes 0, … n - 1 (edges Ei<->(i+1) for i in 0 … n - 2).



The defaults above are in force, so the path is undirected and the probability of moving to the left-hand neighbor is equal to the probability of moving to the right-hand neighbor for each node in the interior.

First, we use a canonical trick and consider immediate neighbors. The access time $H(k, k+1)$ is one less than the return time $R(k+1)$ on a clipped path where node k + 1 is an endpoint.

<Image>

Now we use a fascinating result that I'll prove later when considering the stationary distribution: the return time R(k) for any node is $2m$ / deg(k). Thus

$$H(k, k+1) = R(k) - 1 = 2k / 1 - 1 = 2k + 1.$$

Finally, we proceed inductively to get $H(i, k)$, $0 <= i < k <= n$. We have to reach node k - 1 to get to k; hence

$$H(i, k) = H(i, k - 1) + H(k - 1, k) = H(i, k - 1) + 2k - 1$$

So

$$H(i, k) = H(i, i + 1) + \ldots + H(k - 1, k) = (2i + 1) + (2(i + 1) + 1) + \ldots + (2(k - 1) + 1)$$

Summing,

$$H(i, k) = 2 ((\text{Sum } 1 <= j <= k - 1 \text{ of } 2j + 1) - \text{Sum } 0 <= j <= i - 1 \text{ of } 2j + 1))$$
$$= (k - 1)(k) - (i - 1)(i) + k - i$$
$$= k^2 - i^2.$$

Ordering of Access Times

There is a neat property of access times to the effect that the expected time of touching a set of nodes in order forwards is the same as the expected time of touching these nodes in order backwards. The property is trivial (boils down to the commutativity of addition) in the case of two nodes:

$$K(i, j) = K(j, i) = H(i, j) + H(j, i)$$

But we can extend this to three nodes, via an argument due to Coppersmith, Tetali and Winkler [2]:

$$H(u, v) + H(v, w) + H(w, u) = H(u, w) + H(w, v) + H(v, u)$$

The details of the proof are instructively subtle. We want to show that the expected number of steps for a walk that starts at u, to hit v, then hit w, then hit u again for the first time, is the same as the expected number of steps for a walk that starts at u, to hit w, then hit v, then hit u.

The natural way to do this is to start a walk W at u and end it when it completes a cycle in the forward direction. W has one cycle in the forwards direction and one reverse cycle in the backwards direction. The expected length of the forward cycle is the average of the length of W over all such cyclic walks, weighted by the probabilities of these walks. And the probabilities of each walk and its corresponding reverse walk are the same. To see this, consider a (forwards) walk with outcome $\{u, v1, v2, \ldots vk, u\}$. Its probability is

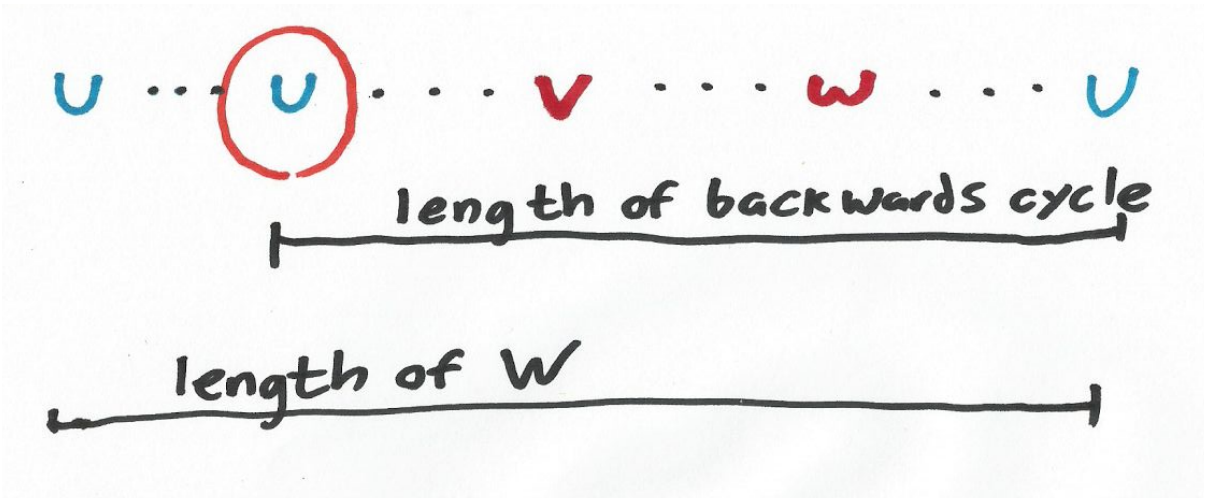
$$Pu\text{->}v1\ Pv1\text{->}v2\ \ldots\ Pvk\text{->}u$$
$$= (1/\deg(u))\ (1/\deg(v1))\ \ldots\ (1/\deg(vk)).$$
$$= (1/\deg(u))\ (1/\deg(vk))\ \ldots\ (1/\deg(v1))$$
$$= Pu\text{->}vk\ Pvk\text{->}vk\text{ -1}\ \ldots\ Pv1\text{->}u$$

Which is the probability of the backwards walk. A sidenote: this is a special case of the **time-reversibility** of a random walk on an undirected graph, a fact whose proof I'll omit:

> **Definition (time-reversibility).** Consider all random walks $\{v0, \ldots vt\}$, where $v0$ is drawn from a probability distribution $D0$. We get a probability distribution Q on the sequences $\{v0, vt\}$. We also get a distribution Dt on the ending node vt. If we reverse each sequence but leave its probability unchanged, we get a new probability distribution Q' on the reversed sequences. Time-reversibility means that Q' is the same distribution we would get considering random walks from Dt.

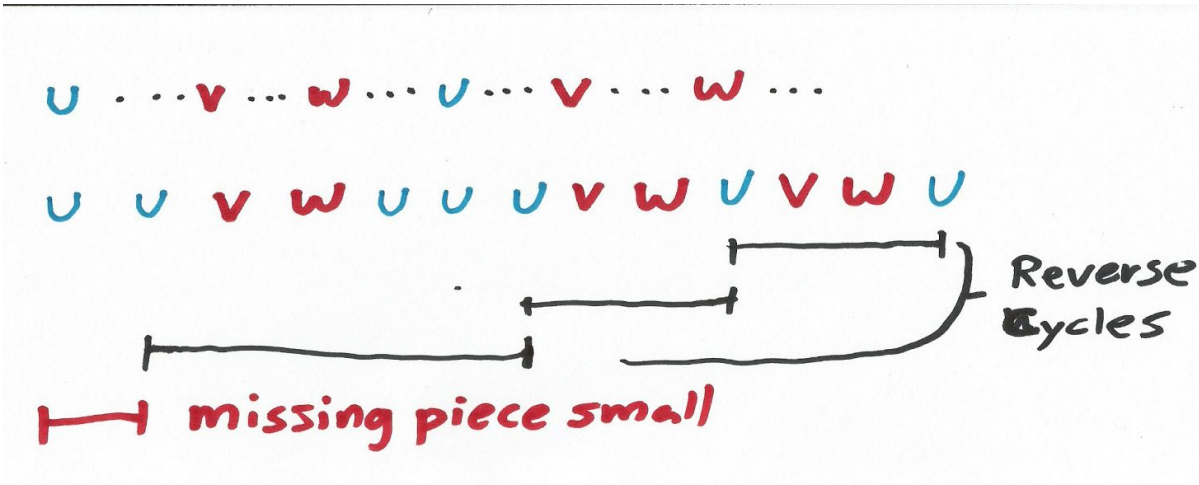
Here we start at u and end at u, so $D0 = Dt$, so, by time reversibility, $Q = Q'$.

So, if we could just say that the length of the backwards cycle is also the length of W, then both the lengths and the probabilities in our weighted average would be the same. But there's an issue. We defined our cycle as starting at u, continuing to v, then to w, and then ending on *the first time* we see u again. But the "first time we see u again" on the reverse walk might not be on the very last node. See below:



To get around this, instead of considering a single cycle, we fix an integer r, start a random walk at u, and end it when we have traversed r cycles. We can show there are also r reverse cycles in the reverse walk. This is because for each cycle in the forward walk, we can find a reverse cycle in the reverse walk, so there are at least r reverse cycles. But for each cycle in the reverse walk, we can find a cycle in the forward walk, so there are no more than r reverse cycles. Hence there are r reverse cycles.

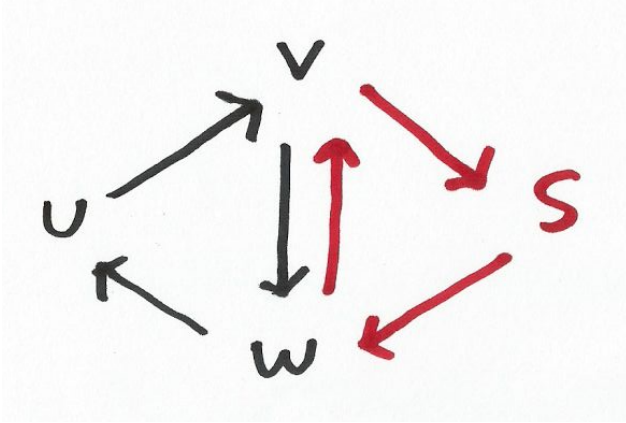
Furthermore, the sum of the forward cycles' length is the length of the walk, and the sum of the backwards cycles is the length of the walk, less the length of a little piece at the beginning, equal to the length of the walk up until the last u before the first v is reached.



As we make r large, this little piece (constant expected length wrt r) becomes insignificant compared to the length of the walk. And so the average length of a forward cycle on the walk approaches the average length of a backwards cycle on the

reverse walk, which shows that the expected lengths of forwards and backwards cycles are the same, which concludes the proof.

An interesting digression--once we have this result for 3-cycles, we can extend it to arbitrary-length cycles, in exactly the same way we can build arbitrary polygons out of triangles. This is in fact a neat graphical representation; you can see how the *v*<->*w* edges might cancel out.



Algebraically, we add

H(u, v) + H(v, w) + H(w, u) = H(u, w) + H(w, v) + H(v, u)

+

H(w, v) + H(v, s) + H(s, w) = H(v, w) + H(w, s) + H(s, v)

And cancel the H(v, w) and H(w, v) that appear on both sides to obtain

H(u, v) + H(v, s) + H(s, w) + H(w, u) = H(u, w) + H(w, s) + H(s,v) + H(v, u).

But we only need the property for 3-cycles to prove the main result of this section: that the nodes of any graph can be ordered so that if u precedes v in the ordering, then H(u, v) <= H(v, u). This ordering can be obtained by fixing any node t and ordering all the nodes u according to the value of H(u, t) - H(t, u).

Proof. If u precedes v, then

H(u, t) - H(t, u) <= H(v, t) - H(t, v), so

H(u, t) + H(t, v) <= H(v, t) + H(t, u).

But subtracting this from

H(u, t) + H(t, v) + H(v, u) = H(u, v) + H(v, t) + H(t, u)

Gives

H(v, u) >= H(u, v) --> H(u, v) <= H(v, u).

How should this be interpreted? I mentioned before that roughly, nodes earlier in the ordering are easy to get away from but hard to reach, and nodes later in the ordering are hard to reach but easy to get away from. This interpretation would be justified if random walks, on average, hit nodes earlier in the list less often than nodes later in the list. I can't prove this in general, but it's easy to prove that, on average, the *first node* in the list is hit less often than the *last node* in the list.

What I'll actually prove is that the return time $R(u)$ for the first node u is larger than the return time $R(v)$ for the last node v, but from arguments in the next section this is equivalent to the above.

All we need is the formula

$$R(s) = (1 / \deg(s)) \text{ Sum}(q \text{ in neighbors}(s) \text{ of } 1 + H(q, s))$$

which you can verify with a moment's thought. Then, since $H(r, u) > H(r, v)$ for all nodes r, by the ordering, each number being averaged in the sum for $R(u)$ is larger than each number being averaged for the sum for $R(v)$. So $R(u) > R(v)$, which at the least sanity-checks our interpretation.

Next up, we verify a far-reaching result with a far clearer interpretation.

The Stationary Distribution: Uniqueness and Existence

In this section, I'll show that, for any connected graph G, there is a unique stationary distribution S on G.

Definition (Stationary Distribution): A probability distribution over nodes $v1 \ldots vk$ of a graph G is stationary if $MS = S$, where S is considered as the column vector $[P(v0), P(v1) \ldots P(vk)]$ and M is the transition probability matrix of G.

I'll also show that a random walk starting from any initial distribution tends to the stationary distribution as the number of steps taken goes to infinity.

Then, in the next section, we'll use it to justify the fundamental assumption of statistical mechanics that all accessible states are equally likely.

Step 1: Existence. Let $S0(v) = \deg(v) / 2m$. $S0(v)$ is a stationary distribution. To see this, let the distribution D be $S0$ at step t, that is, $Dv(t) = S0(v)$. Then

$$Dv(t + 1) = \text{Sum } (i \text{ in neighbors}(v) \text{ of } Di(t) * Pi\text{->}v)$$
$$= \text{Sum } (i \text{ in neighbors}(v) \text{ of } (\deg(i) / 2m) * (1 / \deg(i)))$$
$$= \text{Sum } (i \text{ in neighbors}(v) \text{ of } 1 / 2\text{m})$$
$$= (1 / 2\text{m}) \text{ Sum } (i \text{ in neighbors}(v) \text{ of } 1)$$
$$= \deg(v) / 2m$$
$$= Dv(t)$$

So if a distribution *D* ever equals S0 at any step t, then it remains *D* at step t + 1, which means S0 is stationary.

Step 2: Uniqueness--we prove $S(v) = S0(v)$ for any stationary distribution *S* and each node *v*. We can think of the stationary random walk as a sum of distributions of random walks starting at each node *vi*, weighted by probability $P(vi) = Si$. To see this, decompose the result of the matrix multiplication $S = M\ S$ as follows:

$$S = M\text{^}t\ S = M\text{^}t\ [S1\ 0\ ..\ 0] + M\text{^}t\ [\ 0\ S2\ 0\ \ldots\ 0] + \ldots + M\text{^}t\ [0\ \ldots\ 0\ Sn]$$

Each of these distributions

$$Wi = M\text{^}t\ [0, 0, \ldots, Si, \ldots\ 0, 0]$$

represents the distribution of a walk $Zi = \{vi, \ldots\ vt\}$ starting at node *vi*, after time t.

Now, the catch; each walk *Zi* is guaranteed to encounter node v after some number of steps, since the graph is connected. And after that, the frequency of hitting v is fixed by the well-defined quantity $R(v)$ as $1/\ R(v)$.

So the average probability over all steps that we are at node *v* $Pavg(v)$, which is the total number of steps we are at *v* as a fraction of the total number of steps *t*, tends to the frequency of hitting *v*, $1/\ R(v)$.

Thus the average probability that we are at node *v* approaches $1/\ R(v)$ for each walk *Zi* as *t* grows large. This implies that $Pavg(v)$ in each of the walk distributions Wi approaches $1\ /\ R(v)$. But $S(v)$ is just a weighted sum over $P(v)$ for each *Wi*. So $Savg(v)$, the average value of $S(v)$ over all steps, is a weighted sum over $Pavg(v)$ for each *Wi* (averages are linear). Which means, as *t* goes to infinity,

$$Savg(v) \text{ -> Sum } (Si * Pavg(v) \text{ in } Wi, \text{ over all } Wi)$$
$$= S1\ (1\ /\ R(v)) + S2\ (1\ /\ R(v)) + \ldots + Sn\ (1\ /\ R(v))$$
$$= (S1 + \ldots + Sn)\ (1\ /\ R(v))$$
$$= 1/\ R(v)$$

But $\lim Savg(v) = \lim S(v) = S(v)$ since $S(v)$ is constant with *t*. This means

$$S(v) = 1/\ R(v)$$

This means that $P(v)$ is the same for any stationary distribution *S* and each node v. But since we already have one stationary distribution *S0*, $S = S0$ for any stationary distribution *S*; *S* is unique.

This argument also determines $R(v)$:

$$1/\ R(v) = P(v) \text{ in } S = P(v) \text{ in } S0 = \deg(v)\ /\ 2m$$
$$\rightarrow R(v) = 2m\ /\ \deg(v)$$

As claimed earlier. And we can determine the expected "edge return time" $ER(i\text{->}j)$, the number of steps after an edge has been traversed in one direction until that edge is traversed in the same direction again. Suppose the edge goes from node *i* to node *j*.

Then to traverse the edge again in the same direction, we need to reach node i again. The expected number of steps to do so is

$$2m / \deg(i).$$

But then we need to leave i through the edge from *i* to *j*, which has probability $1/\deg(i)$. The expected number of times we must trial our $1/\deg(i)$ bernoulli variable to get a positive result and leave through the correct edge is $\deg(i)$; hence we expect to reach *i* $\deg(i)$ times before traversing the same edge again, and so

$$ER(i, j) = \deg(i) * 2m / \deg(i) = 2m,$$

Independent of the particular edge.

All Walks Lead to the Stationary Distribution

Next, I'll reprise Lovasz's proof that any distribution tends to the stationary distribution. That is, the probability $Pi->j(t)$ of a node starting at *i* being at *j* after *t* steps tends to $S(j)$. If this is true, then the probability $Pj(t)$ of a node being at *j* after *t* steps tends to $S(j)$ no matter the starting distribution *D0*. We want to use our transition probability matrix *M*, plus the spectral theorem from linear algebra.

> **Theorem (Spectral Theorem [3]):** Any symmetric *n* x *n* matrix has exactly *n* eigenvectors and they can be chosen to form an orthonormal basis for *R^n*.

The problem is that *M* is not symmetric. But we can express it in symmetric form. We know that M = DA, where A is the adjacency matrix of the graph *G* that *M* corresponds to, and *D* is a diagonal matrix where $Dii = 1/\deg(i)$.

> **Definition (Adjacency Matrix) :** *Aij* = 1 if there is an edge *i->j*, 0 otherwise.

You can verify *M* = *DA*, but the idea is that *Mij* = *Aij Dii* so that *Mij* = 0 if there is no edge from *i* to *j*, and otherwise *Mij* is $1/\deg(i)$ as expected.

Now *A* is symmetric (because we assume *G* is undirected, so *Aij* = *Aji*). And multiplying a symmetric matrix on either side by a diagonal matrix preserves symmetry. To see this, note *Nij* = *Aij Dii Djj* = *Aji Djj Dii* = *Nji.* Thus the matrix

$$N = D^{1/2} A D^{1/2}$$

is symmetric.

A quick note: given a diagonal matrix *D* with nonzero entries on the diagonal, we define *D^p* for any number p by the diagonal matrix

$$D^p\ ii = (Dii)^p.$$

This definition satisfies the normal rule of exponentials $D^p D^s = D^(p+s)$, and in particular, $D D^-1 = I$, so that the inverse matrix corresponds to the -1th power. So

$M = DA \rightarrow A = D\text{-}1\ M \rightarrow$

$N = D^{1/2}\ D^\text{-}1\ M\ D^{1/2}$

$= D^{-1/2}\ M\ D^{1/2}$.

Which means

$M = D^{1/2}\ N\ D^{-1/2}$.

Now we have M in terms of the symmetric matrix N. The spectral theorem says we can write

N = Sum(k = 1 to n of $Lk\ Vk\ VkT$)

Where $L1 >= L2 >= \ldots >= Ln$ are the eigenvalues of N and $V1, \ldots Vn$ are the corresponding eigenvectors of unit length, and VkT signifies the transpose of Vk. To see this, recall that a rank n matrix like N, considered as a linear transform, is determined by the values that it sends a set of basis vectors for R^n to. But by the spectral theorem, the eigenvectors Vk are a basis for R^n. So all we must show is that

N Vj = Lk Vj for each Vj.

But this is trivial, because

N Vj = Sum(k = 1 to n of Lk Vk VkT Vj).

Grouping the last terms,

N Vj = Sum(k = 1 to n of Lk Vk (VkT Vk)).

But

VjT Vk = <Vk, Vj>

= 0, k /= j since the Vk's are orthogonal

= 1, k = j since the Vk's are unit length.

Hence

N Vj = Lk Vj.

So the spectral theorem allows this decomposition. Now I have to invoke another result from linear algebra.

> **Theorem (Frobenius-Perron Theorem [4]) :** Let B be a irreducible matrix. (Any matrix that corresponds to a connected undirected graph in the way N does--in the sense that an entry Bij is nonzero iff there is an edge in the graph from i to j--is irreducible). Then
>
> 1. B has a real eigenvalue r that is larger than or equal to all other eigenvalues in modulus.
> 2. r corresponds to a single right eigenvector v.

3. *v* has all positive real components.
4. *v* is the only right eigenvector of *B* with all positive real components.

To use the Frobenius-Perron Theorem, we need to find the eigenvector of *N* with all positive components. But it's easy to see that *Wi* = (deg(*i*))^½ is an eigenvector of *N* with eigenvalue 1, and clearly, all positive components. Hence

*L*1 = 1 > *L*2 > *L*3 > … > *L*n >= -1.

Lovasz also states that *Ln* >= -1 if our graph G is non-bipartite. Bipartite-ness is a particular kind of degeneracy where the nodes of *G* can be partitioned into two sets *U* and *V* such that all edges of *G* join nodes in *U* to nodes in *V*.

Finally, note that the normalized eigenvector *V*1 corresponding to *L*1 is given by

W / (2*m*)^½,

so that *V*1*i* is (deg(*i*) / 2*m*)^½ = (*S*(i)^½).

Now, we're interested in the probability that a random walk, starting at node *i*, is at node *j* after *t* steps. Now this is an entry of *M*^*t*:

Pi->j(*t*) = (*M*^*t*)*ij*.

To see this, consider the initial distribution D0 corresponding to a walk starting at i:

D0i = 1, *D0j* = 0, *j* /= *i*

Hit this distribution with *M t* times to produce *Dt*. Now the catch: matrix multiplication is associative, so

Dt = (*M*^*t*) *D0*.

But

Dtk = ((*M*^*t*) *D0*)*k* = (*M*^*t*)*ik* so *Dtj* = *Pi->j*(*t*) = (*M*^*t*)*ij*.

So we want to figure out the limiting value of *M*^*t*. We have

M^*t* = *D*^½ *N*^*t D*^-½

Since when multiplying out the formula for M in terms of N, the inner diagonal pieces cancel off.

Let's take our spectral formula for *N* and square it. When we do this, we get a bunch of terms of the form

LkVkVkT * *LjVjVjT*
= *LkLj VkVkT VjVjT* = *LkLj Vk* (*VkT Vj*) *VjT*
= *LkLj Vk* (*VkT Vj*) *VjT*
= *LkLj* <*Vk*, *Vj*> *VjT*.

Recall that the central inner product is 1 for k = j, 0 otherwise. So all the terms with k /= j disappear and we are left with

N^2 = Sum(k = 1 to n of $Lk^2\ Vk\ VkT$)

$\rightarrow N^t$ = Sum(k = 1 to n of $Lk^t\ Vk\ VkT$)

Which means

M^t = Sum(k = 1 to n of $Lk^t\ D^{\frac{1}{2}}\ Vk\ VkT\ D^{-\frac{1}{2}}$)

= Q + Sum(k = 2 to n of $Lk^t\ D^{\frac{1}{2}}\ Vk\ VkT\ D^{-\frac{1}{2}}$)

Where

$Q = D^{\frac{1}{2}}\ V1\ V1T\ D^{-\frac{1}{2}}$

But if G is not bipartite then $|Lk| < 1$ for k = 2, … n, and so $M^t \rightarrow Q$ as t goes to infinity. So $(M^t)ij \rightarrow Qij$ as t goes to infinity. That is, the probability that a random walk starting at i is at j after t steps goes to Qij as t goes to infinity. What is Qij? It is

$(D^{\frac{1}{2}})ii\ V1i\ V1j\ D^{-\frac{1}{2}}$

= $(1 / \deg(i))^{\frac{1}{2}}\ (\deg(i) / 2m)^{\frac{1}{2}}\ (\deg(j) / 2m)^{\frac{1}{2}}\ (\deg(j)^{\frac{1}{2}}$)

= $\deg(j) / 2m$

= $S(j)$.

So the probability of a random walk starting at any node being at j after t steps tends to $S(j)$, which is what we wanted to prove. Next, we'll apply this result to statistical mechanics.

Fundamental Assumption of Statistical Mechanics

First, a very quick bit of background on statistical mechanics. Because of quantum mechanics, most physical systems can be thought of as having a finite number of states. So we can consider a graph whose nodes are the states. For example, in the simple Einstein model, a solid is viewed as a lattice of atoms, each of which contains a certain discrete positive amount of energy quanta that can be transferred with neighbors. The state of this solid is a list of the number of quanta contained in each atom. Typically, we constrain the set of allowed states by a set of macroscopic system variables--in the case of the Einstein solid, the macrostate is the total number of energy quanta.

The edges of the graph, which represent transitions of the system, depend on the system. In the Einstein solid state graph, there is an edge between every pair of states $S1$, $S2$, such that across the states, the number of quanta in all but two

neighboring atoms i and j are the same, but in *S1* quanta(i) = x, quanta(j) = y whereas in *S2* quanta(i) = x + 1, quanta(i) = y - 1. This corresponds to atom *i* gaining a quantum of energy from atom *j.*

A random walk on a graph corresponding to a physical system corresponds to the time-evolution of that system. To make this correspondence accurate, we have to assume that the physical system evolves in discrete time steps in place of a more nuanced probability density model of transitioning in time. But we don't lose any information this way about the limiting likelihoods of the states. If we want more information about when transitions happen, we can generate a series of time values from a Poisson distribution:

$t0 = 0$, $ti = t(i\text{-}1)$ + Poi(L), L average time of transition

And map each *ti* to a corresponding node in the outcome of a random walk. But this is not our main focus.

Now I'll check the assumptions we made about general graphs above and see how these assumptions relate to graphs representing physical systems. The assumption of connectedness carries over--we want to consider a set of states such that every state can be reached from every other. The assumption that $Pi\text{->}j \sim \deg(i) = \deg(i) / 2m$ does not necessarily hold--I'll return to this later. The assumption of undirectedness still holds. In general, the microscopic laws of physics are time-reversible, so any transition that occurs in one direction can occur in the other.

In fact, we're going to assume--the key assumption of this section:

Definition (Strong Reversibility): $Pi\text{->}j = Pj\text{->}i$.

I originally thought--when drafting this paper--strong reversibility made physical sense. But I realized in the course of research it is generally not true; it is true, according to Lewis [5], only if the fundamental assumption of statistical mechanics, which I had hoped to derive, is true.

Definition (Fundamental Assumption of Statistical Mechanics): All possible states tend to equal likelihood with time.

If the system has reached equilibrium, then a general principle, argued by Lewis, forbidding "cyclic equilibrium" comes into effect. That is, consider a large number (much larger than the number of possible states) of copies of our system *S1*… *Sk*, and call *Ni*, *Nj* the average number of such systems in state *i* and state *j*. By the nonexistence of cyclic equilibrium as argued by Lewis, it is not physically possible that Ni and Nj remain constant while the rate of transition from *i* to *j* is different from the rate of transition from j to i:

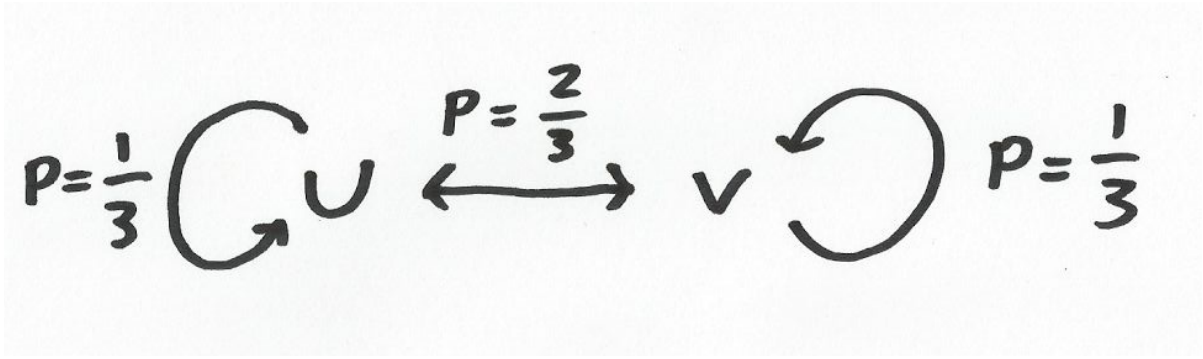
$Ni\ Pi\text{->}j \ /= Nj\ Pj\text{->}i.$

Hence in equilibrium,

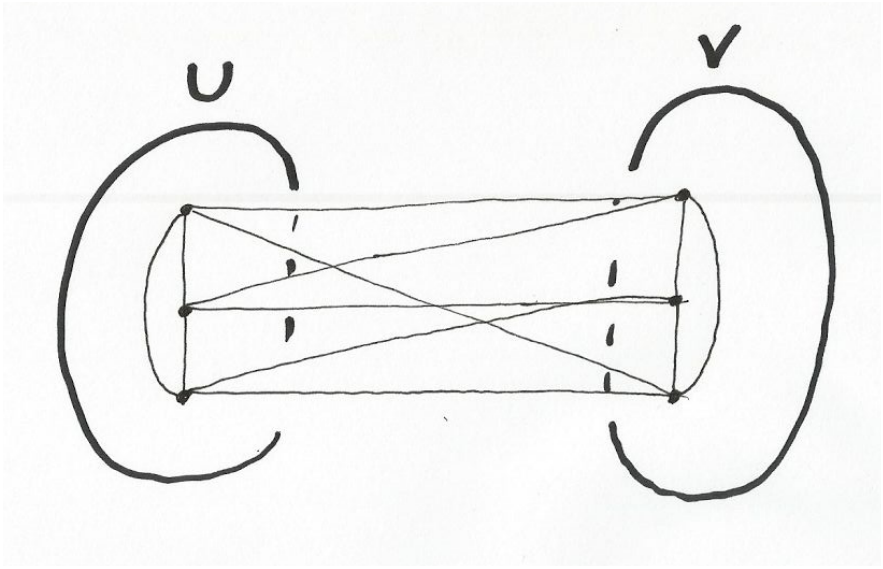
$$Pi\text{->}j = Pj\text{->}i \text{ only if } Ni = Nj,$$

Which condition is the fundamental assumption of statistical mechanics, since Ni, Nj = $S(i)$, $S(j)$ in equilibrium.

But I'll argue the converse anyway. If a physical system is strongly reversible, I can construct a graph that models whatever transition probabilities $Pi\text{<->}j$ that system has, to an arbitrary precision. For example, I can model a system with this graph T:



This system, as drawn, violates the assumption we made in previous sections that $Pi\text{->}j$ = deg(i). So we can't quote the proof from above that all distributions tend to S. So I'll replace it with this graph T':



T' is regular of degree 3 and $Pn\text{->}m$ = 1/ deg(n) = 1/3 for all n, m that are connected by an edge. But if we think of the nodes in each bubble as a single node, this graph mimics the behavior of the graph T.

I can do this in general. Given a graph G on vertices $v1$, …, vn with given transition probabilities Pi->j that are rational with denominator N (N some integer), I'll replace it with a graph G', with $n * N$ nodes $s1$, $s2$, …, $s(n * N)$. To do this, I replace

each node *vi* of *G* with *N* nodes. Define, for each *i*, *Qi* = {*qi1*, … *qiN*}, the set of *N* nodes in *G'* corresponding to node *vi* in *G*. Each node in *Qi* is connected to *N Pi->j* of the nodes (it doesn't matter which) in *Qj*, so that the probability of going to a node in *Qj* from a node in *Qi* is *N Pi->i / N = Pi->j.* Hence the graph *G'* thus produced mimics the graph *G*, in the precise sense that *P*(*Qj->Qi*) = *Pi->j*.

In terms of random walks, the probability that a walk *W* starting from initial distribution *D0* = [*P*(*v*1), …, *P*(*vn*)] on *G* is at node *i* after *t* steps is exactly the same as the probability that a walk *V* is at a node in *Qi* after t steps, starting from any distribution *D* on *G'* where the sum *P*(*qj*) over nodes in *Qi* = *P*(*vi*).

Now *G*' is subject to the theorem that all distributions tend to the stationary distribution. That stationary distribution on *G'* is uniform, since *G'* is regular:

$$P(si) = P = N / 2m = 1 / (N * n).$$

So the probability of a random walk in *G'* being in each set *Qi* tends to 1 / *n*.

Which implies that the probability of a random walk in *G* being in node *vi* tends to 1 / *n*.

Returning to our interpretation of the nodes of *G* as states of a physical system, this means that all possible states tend to the equal likelihood 1 / *n* with increasing time. So we have verified the fundamental assumption of statistical mechanics for strongly reversible physical systems.

Conclusion

In writing this paper I gained tremendous appreciation for the generality of the concepts "graph" and "random walk", both in application and in relation to other mathematical domains.

We've gotten a glimpse of the relationship with linear algebra, but random walks on graphs also have close ties to topics in complex analysis like harmonicity. Briefly, we say that the function *P*: *V* -> *R* on the set of vertices *V* of a graph is discrete harmonic if it satisfies the "mean value property" that for each node *v* of the graph,

$$P(v) = (1 / \deg(v)) \text{ Sum}(u \text{ in neighbors}(v), P(u)).$$

We can define discrete harmonic functions to model voltage in an electrical network or the equilibrium positions of a set of springs. Or, the expected end position of a random walk. Suppose we have a subset *S* of the vertex set *V* and and a function *P0*: *S* -> *R*.

Then we define *P*(*v*) for *v* in *V* as the expectation of *P0*(*s*), where *s* is the random node where a random walk starting at *v* first hits *s*. *P*(*v*) is discrete harmonic--because

expected values satisfy the mean value property above. And $P(u) = P0(u)$ for u in S. P is closely related to the probabilistic interpretation of harmonic measure. To see this, approximate the unit disk by a planar graph that fills out the disk--maybe a rectangular lattice. The boundary of the disk is analogous to the set of outermost nodes of the lattice S. Define $P0(v) = 1$ for v in some "contiguous" subset C of S (analogous to the curve being measured) and $P0(v) = 0$ for v in $S - C$. Then $P(v)$ inside the disk is a discrete harmonic function that agrees with $P0$ on the boundary, analogous to the solution P' of the corresponding Dirichlet problem on the unit disk. But $P(v)$ is the expectation that a walk starting at v ends on C. So by analogy, we might expect that $P'(z)$, the harmonic measure of C from z, represents the probability that a "random walk" on the unit disk starting at z ends on C.

And the applications Lovasz surveys range from distributed computing to electrical networks to estimating volumes of solid bodies, not to mention statistical mechanics. Suppose that you have to estimate the volume of a body B in R^n. We can do this if we can generate a random element of a superset S of B of known volume and count how many times we hit B. To do this (essentially; the details are complicated) we fill out S with a fine regular graph G and start a random walk on G. Our theorem about the stationary distribution ensures that if we walk long enough, the place we end up at is essentially uniformly distributed. To do this in practice requires estimates on how fast distributions tend to the stationary distribution, which is another ballgame entirely.

Random walks on graphs have so many rich applications and mathematical connections that, to give you a uniformly distributed and representative sample of the field, I would have to walk, or write, a long time. But I hope that in this short paper I've conveyed a little of the sense of power and generality and connectedness, of the study of random walks on graphs.

References

1. L. Lovasz, Random Walks on Graphs, 'Combinatorics, Paul Erdos is Eighty', 2(1993), 1-46.
2. D. Coppersmith, P. Tetali, P. Winkler, Collisions Among Random Walks on a Graph, 'SIAM Journal on Discrete Mathematics', 6, 3(1993), 363-374
3. L. El Ghaoui, Hyper-Textbook: Optimization Models and Applications, published online (2014). https://inst.eecs.berkeley.edu/~ee127a/book/login/index.html
4. Wikipedia, 'Perron-Frobenius Theorem', retrieved June 2015. https://en.wikipedia.org/wiki/Perron–Frobenius_theorem
5. N. Lewis, A New Principle of Equilibrium, 'Proceedings of the National Academy of Sciences of the United States of America', 3, 11(1925), 179-183.