Global Information from Local Observation

Itai Benjamini and László Lovász
Microsoft Research, One Microsoft Way, Redmond, WA 98052
itai@wisdom.weizmann.ac.il, lovasz@microsoft.com

Abstract

We observe a certain random process on a graph "locally", i.e., in the neighborhood of a node, and would like to derive information about "global" properties of the graph. For example, what can we know about a graph based on observing the returns of a random walk to a given node?

Our main result concerns a graph embedded in an orientable surface with genus g, and a process, consisting of random excitations of edges and random balancing around nodes and faces. It is shown how to obtain the genus of the surface in polynomial time from local observations of the process restricted to a connected subgraph whose size is (essentially) $O(g^2)$.

1 Introduction

At least since Pólya's observation (1921) that "a drank man will return home while a drank bird might lose its way forever", it is known that geometric properties of the underling space can manifest themselves in the behavior of a random process taking place on the space. More recently, random processes were used to retrieve information about the underlying space. E.g. sampling, volume estimates of convex bodies or scenery reconstruction along random walk paths (see for instance [11], [2] and [9]).

Several examples can be found in physics: consider a stationary spin system on a graph, such as Glauber dynamics for the Ising model. What properties of the graph can be inferred from properties of the process (for instance, are there interesting relations between the spectrum of the Ising dynamics and the graph spectrum)?

We may ask an even harder question: rather then letting a random walker wonder around and gather information, or observe the behavior of the Ising model, suppose that we can observe the random process only locally, say in a fixed bounded neighborhood of a node. Can non-trivial global observation be distilled from such *local* observations? Can we put reasonable restriction (say, polynomiality) on the observation time?

As a simple example, suppose that we start a random walk on a graph, but we can only observe it locally, at a single node. What properties of the graph can be inferred from this? Even to this simple question, we don't know the answer.

Facing this harder challenge, one might start with devising custom-made variants of the standard random processes (spin systems, random walks), which can be analyzed, and provide ways to compute global invariants using natural "physical" systems.

Below we will study a reasonably simple process, the "Noisy Circulator" with local operations, living on the edges of a graph embedded in an orientable surface. The process consists of adding mass 1 to randomly chosen edges at a slow rate, and balancing the flow into vertices or around faces at random, at a faster rate (details below). We know a priori bounds on the size of the graph and on the genus. We will show how to extract, almost surely and in polynomial time in the size of the graph, the exact genus of the surface by observing the restriction of the process on a "small" connected subgraph.

Although this construction might be useful and have some advantages even from a pure algorithmic view point, our goal is not to devise a distributed algorithm for finding the genus, but to provide a non-trivial example for the "global information from local observation" phenomenon.

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

The analysis of the algorithm, which involves a topological theorem and a statistical element, has some independent interest. In the topological part, we define *rotation-free circulations* (that are also circulations on the dual map). These can be considered as discrete analogues of analytic functions. Every homology class of circulations contains exactly one rotation-free circulation, so the dimension of their space can be used to find the genus of the surface. A key result is a discrete analogue of the fact that an analytic function is determined by its values on an arbitrarily small neighborhood of a point: it asserts that every connected piece of the vanishing set of a rotation-free circulation can be separated from the rest of the graph by a small number of nodes.

We briefly discuss some other examples where global information can be obtained from local observation (these results are less complete).

2 Main result: randomized circulations

The Graph. Let S be a closed compact surface, and consider a *map* on S, i.e., a graph $G = (V, E)$ embedded in S so that each face is a disc. We can describe the map as a triple $G = (V, E, \mathcal{F})$, where V is the set of nodes, E is the set of edges, and $\mathcal{F}$ is the set of faces of G. We fix a reference orientation of G; then each edge $e \in E$ has a *tail* $t(e) \in V$, a *head* $h(e) \in V$, a *right shore* $r(e) \in \mathcal{F}$, and a *left shore* $l(e) \in \mathcal{F}$.

The embedding of G defines a *dual map* G^*. Combinatorially, we can think of G^* as the triple $(\mathcal{F}, E, V)$, where the meaning of "node" and "face", "head" and "right shore", and "tail" and "left shore" is interchanged.

For each node v, let $\delta v \in \mathbb{R}^E$ denote the coboundary of v:

$$(\delta v)_e = \begin{cases} 1, & \text{if } h(e) = v, \\ -1, & \text{if } t(e) = v, \\ 0, & \text{otherwise.} \end{cases}$$

Loops attached at v have a 0 coordinate (their heads and tails cancel out each other). If there are no loops at v, then $|\delta v|^2 = d_v$ is the degree of v.

For every face $F \in \mathcal{F}$, we denote by $\partial F \in \mathbb{R}^E$ the boundary of F:

$$(\partial F)_e = \begin{cases} 1, & \text{if } r(e) = F, \\ -1, & \text{if } l(e) = F, \\ 0, & \text{otherwise.} \end{cases}$$

If the face F is on both shores of an edge, then this edge has 0 coordinate in ∂F: walking along the boundary we traverse the edge twice in different directions (since the surface is orientable). If there are no such edges, then $d_F = |\partial F|^2$ is the length of the cycle bounding F.

The Noisy Circulator. Let $p > 0$ be fixed. We start with the vector $x = 0 \in \mathbb{R}^E$. At each step, the following two operations are carried out on the current vector $x \in \mathbb{R}^E$:

(a) [Node balancing.] We choose a random node v, and subtract from x the vector $(x^{\mathsf{T}}(\delta v)/d_v)\delta v$.

(b) [Face balancing.] We choose a random face F, and subtract from x the vector $(x^{\mathsf{T}}(\partial F)/d_F)\partial F$.

In addition, with the given probability $p > 0$, we do the following:

(c) [Edge excitation.] We choose a random edge e, and add 1 to x_e.

Immediately after a node balancing step, the node v just balanced satisfies the flow condition; a subsequent other node balancing may destroy this. Similarly, after a face balancing step, the net "rotation" around the newly balanced face is 0, but this may be destroyed when other faces are balanced. However, we'll see that under repeated application of (a) and (b), any vector converges to a *rotation-free circulation*, i.e., to a circulation with 0 rotation around each face.

We'll also prove that rotation-free circulations form a linear space $\mathcal{C}$ with $\dim(\mathcal{C}) = 2g$, so to recover the genus g it suffices to determine this dimension.

Genus estimate algorithm: outline. Let $U \subseteq V(G)$ induce a connected subgraph of G. Let E_0 be the set of edges incident with U; this set is our "observation window". We say the U is *well-connected* if every set of nodes that separates U from any cycle that is not contractible in S, has at least $4g$ elements. Note that if the graph is embedded "reasonably uniformly" in the surface so that the ball

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

with radius g about each node is planar and is (say) a grid, then we can take such a ball as our observation window. The well-connectedness condition as defined above is necessary at least for our algorithm to work (see remark 7 below).

Let $x(t) \in \mathbb{R}^E$ be the vector of edge-weights after t steps, and let $y(t)$ be the restriction of $x(t) \in \mathbb{R}^{E_0}$ to the edges in E_0. So we can observe the sequence random vectors $y(0), y(1), \ldots$.

Let $x'(t)$ be the projection of $x(t)$ onto $\mathcal{C}$, and let $y'(t)$ be the restriction of $x'(t)$ to the edges in E_0. Because of the excitation steps (c), after a sufficiently long time, the vectors $x'(0), \ldots, x'(t-1)$ will span the space $\mathcal{C}$. So the rank of this set of vectors gives us the dimension of $\mathcal{C}$, and so by Theorem 4, it gives us the genus of the surface. Our key topological result will assert that if U is well-connected, then the restriction to E_0 is one-to-one on $\mathcal{C}$ (Theorem 6), and so this rank is the same as the rank of the set $\{y'(0), y'(1), \ldots, y'(t)\}$.

Unfortunately, we cannot observe the vectors $y'(t)$, only the vectors $y(t)$. But (at least if p is small) we expect the errors $y''(t) = y(t) - y'(t)$ to be small, at least as long as no excitation step (c) occurs. The speed of convergence depends on the eigenvalue gap of the transition matrix of the (undirected) random walk on G and G^*.

So we are lead to the following quite standard statistical problem: there is a sequence $y'(0), y'(1), \ldots$ of vectors in $\mathbb{R}^k$, which span a linear subspace $\mathcal{L}$. We observe the sequence $y(t) = y'(t) + y''(t)$, where the "error" $y''(t)$ is small on the average. We want to find the dimension of $\mathcal{L}$.

The random vectors $x(t)$ or $x'(t)$ are not independent (in the probabilistic sense); but the differences $x'(t+1) - x'(t)$ are. Indeed, node balancing and face balancing don't change x'; so if no excitation occurred in step $t+1$, then $x'(t+1) - x'(t) = 0$, and if new flow was created on edge e, then it depends only on e. Hence the vectors $y'(t+1) - y'(t)$ are also mutually independent.

Since $y'(t+1) - y'(t) = 0$ most of the time, it makes sense to aggregate N of these terms to one. So we consider the vectors $z(t) = y(Nt) - y(N(t-1))$ and $z'(t) = y'(Nt) - y'(N(t-1))$. Then the vectors $z'(t)$ are mutually independent samples from some distribution on L. (The errors $z''(t) = z(t) - z'(t)$ may be dependent.)

Genus estimate algorithm: formal description. We assume that we are given the (small) excitation probability p, and upper bounds $n_0 \geq n + m + f$ and $g_0 \geq g$. Let m_0 be the number of edges incident with U. Define the parameters

$$\varepsilon = \frac{1}{10} n_0^{-m_0 n_0}, \quad T = 10^{10} n_0^2 g_0^2,$$
$$T' = 160 g_0, \qquad N = \left\lceil \frac{1}{10^6 n_0 g_0 p} \right\rceil.$$

Let $y(t) \in \mathbb{R}^{E_0}$ be the vector we observe in our window after t steps, and let $z(t) = y(Nt) - y(N(t-1))$.

Construct a sequence of integers $t_1, t_2, \ldots, \in [0, T-1]$ as follows. If we have $t_1, t_2, \ldots, t_k$, then compute the linear hull $\mathcal{L}(k)$ of $z(t_1), z(t_2), \ldots, z(t_k)$. Let $H(k)$ be the set of integers $t \in [0, T-1]$ for which the vector $z(t)$ is farther from $\mathcal{L}(k)$ than ε. If $|H(k)| < T'$, then return $k/2$ as your guess for g. Else, choose a number $t_{k+1} \in H(k)$ randomly and uniformly. If k becomes larger that $2g_0$, declare the procedure a failure and stop.

Analysis of the algorithm. The main result of this paper is the following.

Theorem 1 *Suppose that U is well-connected and $p < 1/(10^{13} g_0^3 m_0 n_0^7 \ln n_0)$. Then the Genus Estimate Algorithm returns the correct genus with probability at least $4/5$. The algorithm runs in time $O(g_0 n_0^4 / p)$.*

If you find that a success probability of 4/5 is not reassuring enough, independent repetition of the observation can boost this arbitrarily close to 1.

3 Properties of rotation-free circulations

3.1 Circulations and homology

A vector $\phi \in \mathbb{R}^E$ is a *circulation* if

$$\phi \cdot \delta v = \sum_{e:\ h(e)=v} \phi(e) - \sum_{e:\ t(e)=v} \phi(e) = 0.$$

Each vector ∂F is a circulation; circulations that are linear combinations of vectors ∂F are called

null-homologous. Two circulations ϕ and ϕ' are called *homologous* if $\phi - \phi'$ is null-homologous.

Let ϕ be a circulation on G. We say that ϕ is *rotation-free* if for every face $F \in \mathcal{F}$, we have

$$\phi \cdot \partial F = 0.$$

This is equivalent to saying that ϕ is a circulation on the dual map G^*.

Remark 2 Rotation-free circulations are closely related to *discrete analytic functions* and are essentially the same as *discrete holomorphic 1-forms.* These functions were introduced for the case of the square grid a long time ago [7, 6]. For the case of a general planar graph, the notion is implicit in [3]. For a detailed treatment see [12].

To explain the connection, let ϕ be a rotation-free circulation on a graph G embedded in a surface. Consider a planar piece of the surface. Then on the set $\mathcal{F}'$ of faces contained in this planar piece, we have a function $\sigma : \mathcal{F}' \to \mathbb{R}$ such that $\partial\sigma = \phi$, i.e., $\phi(e) = \sigma(r(e)) - \sigma(l(e))$ for every edge e. Similarly, we have a function $\pi : V' \to \mathbb{R}$ (where V' is the set of nodes in this planar piece), such that $\delta\pi = \phi$, i.e., $\phi(e) = \pi(t(e)) - \pi(h(e))$ for every edge e. We can think of π and σ as the real and imaginary parts of a (discrete) analytic function. The relation $\delta\pi = \rho\phi$ is then a discrete analogue of the Cauchy–Riemann equations.

Thus we have the two orthogonal linear subspaces: $\mathcal{A} \subseteq \mathbb{R}^E$ generated by the vectors δv ($v \in V$) and $\mathcal{B} \subseteq \mathbb{R}^E$ generated by the vectors ∂F ($F \in \mathcal{F}$). Vectors in $\mathcal{B}$ are 0-homologous circulations. The orthogonal complement $\mathcal{A}^\perp$ is the space of all circulations, and $\mathcal{B}^\perp$ is the space of circulations on the dual graph. The intersection $\mathcal{C} = \mathcal{A}^\perp \cap \mathcal{B}^\perp$ is the space of rotation-free circulations. So $\mathbb{R}^E = \mathcal{A} \oplus \mathcal{B} \oplus \mathcal{C}$. From this picture we conclude the following.

Lemma 3 *Every circulation is homologous to a unique rotation-free circulation.*

We defined $\mathcal{A}$ by n generators, which are not independent: their sum is 0. But it is easy to see that this is the only linear relation among them, so $\dim\mathcal{A} = n-1$. Similarly, $\dim\mathcal{B} = f-1$, and hence by Euler's Formula,

$$\dim\mathcal{C} = m-(n-1)-(f-1) = m-n-f+2 = 2g.$$

Thus we have proved:

Theorem 4 *The dimension of the space $\mathcal{C}$ of rotation-free circulations is $2g$.*

(This would also follow from Lemma 3, which implies that $\mathcal{C}$ is isomorphic to the first homology group of S over the reals.)

3.2 Nonvanishing of rotation-free circulations

We start with a simple lemma about maps. For every face F, let a_F denote the number of times the orientation changes if we move along the the boundary of F. For every node v, let b_v denote the number of times the orientation changes in their cyclic order as they emanate from v.

Lemma 5 *Let $G = (V, E, \mathcal{F})$ be any digraph embedded on an orientable surface S of genus g. Then*

$$\sum_{F\in\mathcal{F}} (a_F - 2) + \sum_{v\in V} (b_v - 2) = 4g - 4.$$

Proof. Clearly

$$\sum_F a_F = \sum_v (d_v - b_v),$$

and so, using Euler's formula,

$$\sum_F a_F + \sum_v b_v = \sum_v d_v = 2m = 2n + 2f + 4g - 4.$$

Rearranging and dividing by 2, we get the equality in the lemma. □

Theorem 6 *Let G be a graph embedded in an orientable surface S of genus $g > 0$ so that all faces are discs. Let ϕ be a non-zero rotation-free circulation on G and let G' be the subgraph of G on which ϕ does not vanish. Suppose that ϕ vanishes on all edges incident with a connected subgraph U of G. Then U can be separated from G' by at most $4g-3$ nodes.*

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

Remark 7 The assumption that the connectivity between U and the rest of the graph must be linear in g is sharp in the following sense. Suppose X is a connected induced subgraph of G separated from the rest of G by $\leq 2g$ nodes, and suppose (for simplicity) that the subgraph induced by X and its neighborhood is embedded in a subset of S that is topologically a disc. Contract X to a single node x, and erase the resulting multiplicities of edges. We get a graph G' still embedded in S so that each face is a disc. Thus this graph has a $(2g)$-dimensional space of circulations, and hence there is a non-zero rotation-free circulation ψ vanishing on $2g-1$ of the edges incident with x. Since this is a circulation, it must vanish on all the edges incident with x. Uncontracting X, and extending ψ with 0-s to the edges of X, it is not hard to check that we get a rotation-free circulation.

Proof. Let W be the connected component of $G \setminus V(G')$ containing U, and let Y denote the set of nodes in $V(G) \setminus V(W)$ adjacent to W.

Consider an edge e with $\phi(e) = 0$. If e is not a loop, then we can contract e and get a map on the same surface with a rotation-free flow on it. If $G-e$ is still a map, i.e., every face is a disc, then ϕ is a rotation-free flow on it. If $G-e$ is not a map, then both sides of e must be the same face. So we can eliminate edges with $\phi(e) = 0$ unless $h(e) = t(e)$ and $r(e) = l(e)$ (we call these edges strange loops). In this latter case, we can change $\phi(e)$ to any non-zero value and still have a rotation-free flow.

Applying this reduction procedure, we may assume that $W = \{w\}$ consists of a single node, and the only edges with $\phi = 0$ are the edges between w and Y, or between two nodes of Y. We cannot try to contract edges between nodes in Y (we don't want to reduce the size of Y), but we can try to delete them; if this does not work, then every such edge must have $r(e) = l(e)$.

Also, if more than one edge remains between w and a node $y \in Y$, then each of them has $r(e) = l(e)$ (else, one of them could be deleted). Note that we may have some strange loops attached at w. Let D be the number of edges between w and Y.

Re-orient each edge with $\phi \neq 0$ in the direction of the flow ϕ, and orient the edges between w and Y alternatingly in an out from w. Orient the edges with $\phi = 0$ between two nodes of Y arbitrarily. We get a digraph G_1.

It is easy to check that G_1 has no sources or sinks, so $b_v \geq 2$ for every node v, and of course $b_w \geq |Y|-1$. Furthermore, every face either has an edge with $\phi > 0$ on its boundary, or an edge with $r(e) = l(e)$. If a face has at least one edge with $\phi > 0$, then it cannot be bounded by a directed cycle, since ϕ would add up to a positive number on its boundary. If a face boundary goes through an edge with $r(e) = l(e)$, then it goes through it twice in different directions, so again it is not directed. So we have $a_F \geq 2$ for every face.

Substituting in Lemma 5, we get that $|Y|-1 \leq 4g-4$, or $|Y| \leq d_w \leq 4g-3$. Since Y separates U from G', this proves the theorem. □

4 Proof of Theorem 1

4.1 Bounding the error

It will be convenient to introduce two further parameters:

$$\delta = n_0^{-4g_0m_0n_0}, \qquad a = \lfloor \frac{1}{10^{12}g_0^2n_0^2p} \rfloor.$$

Consider the matrices

$$A = \frac{1}{n}\sum_{v \in V} \frac{1}{d_v}(\delta v)(\delta v)^{\mathsf{T}},$$

and

$$B = \frac{1}{f}\sum_{F \in \mathcal{F}} \frac{1}{d_F}(\partial F)(\partial F)^{\mathsf{T}}.$$

Let λ_1 and λ_2 be the smallest positive eigenvalue of A and B, respectively, and let $\mu = \min\{\lambda_1, \lambda_2\}$.

We can write $A = MM^{\mathsf{T}}$, where M is a $E \times V$ matrix defined by

$$M_{iv} = \begin{cases} \frac{1}{\sqrt{nd_v}}, & \text{if } h(i) = v, \\ \frac{-1}{\sqrt{nd_v}}, & \text{if } t(i) = v, \\ 0, & \text{otherwise.} \end{cases}$$

Then A has the same nonzero eigenvalues as the matrix $M^{\mathsf{T}}M$, which is a $V \times V$ matrix with

$$(M^{\mathsf{T}}M)_{uv} = \begin{cases} \frac{-1}{n\sqrt{d_ud_v}}, & \text{if } uv \in E, \\ \frac{1}{n}, & \text{if } u = v, \\ 0, & \text{otherwise.} \end{cases}$$

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

This is $(1/n)$ times the symmetrized form of the transition matrix of the random walk, from which it follows by well-known results that $\lambda_1 \geq n^{-4}$. Similarly, $\lambda_2 \geq f^{-4}$, so $\mu \geq \min\{n^{-4}, f^{-4}\} > n_0^{-4}$.

Let $T_t : \mathbb{R}^E \to \mathbb{R}^E$ denote the (random) linear mapping that (a) and (b) generate in step t. Note that the subspace $\mathcal{C}$ is invariant under T_t. Let $W(t,s) = T_s T_{s-1} \dots T_{t+1}$,

$$u(t) = \begin{cases} e_j, & \text{if in step } t \text{ edge } j \text{ was excited,} \\ 0, & \text{otherwise.} \end{cases}$$

and $u(t,s) = W(t,s)u(t)$. Hence

$$x(s) = \sum_{t=0}^{s} u(t,s).$$

Let $u_1(t,s)$, $u_2(t,s)$ and $u_3(t,s)$ denote the orthogonal projections of $u(t,s)$ to $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, respectively, and let $u_i(t) = u_i(t,t)$. This notation is somewhat redundant, since $\mathcal{C}$ is invariant under T_t, and hence $u_3(t,s) = u_3(t)$ for every $s \geq t$. The "error part" is $w(t,s) = u_1(t,s) + u_2(t,s)$. Thus

$$x'(s) = \sum_{t=0}^{s} u_3(t), \qquad x''(s) = \sum_{t=0}^{s} w(t,s).$$

This x'' in turn can be split into "old" and "new" errors:

$$x''(s) = \sum_{t=0}^{s-a} + \sum_{t=s-a+1}^{s} = X_1(s) + X_2(s).$$

Note that $X_2(s) = 0$ unless an excitation occurs in the interval $[s-a+1, s]$.

We also need a simple lemma. Let $a_1, \dots, a_k \in \mathbb{R}^n$, and let $A = (1/k)\sum_{i=1}^{k} a_i a_i^{\mathsf{T}}$. Let λ be the smallest positive eigenvalue of A. Let L be the linear subspace generated by $a_1, \dots, a_k$, and $L^{\perp}$, its orthogonal complement.

For $x \in \mathbb{R}^n$, define a Markov chain $X^0, X^1, \dots \in \mathbb{R}^n$ as follows: start with $X^0 = x$. Given X^t, choose a vector a_i (uniformly and randomly), and let

$$X^{t+1} = X^t - \frac{a_i^{\mathsf{T}} X^t}{a_i^{\mathsf{T}} a_i} a_i.$$

Let y and z be the orthogonal projections of x onto $L^{\perp}$ and L, respectively. Then $X^t \to y$, and in fact the following lemma (whose proof is omitted) describes the rate of convergence:

Lemma 8 $\mathsf{E}\left(|X^t - y|^2\right) \leq (1-\lambda)^t |z|^2$.

From this we derive:

Lemma 9 $\mathsf{E}(|X_1(s)|^2) \leq 5pn_0^4 e^{-a/n_0^4}$.

Proof. Let us fix the "excitations" $u(0), u(1), \dots$, and let E_u denote expectation conditional on these. Lemma 8 implies that

$$\mathsf{E}_u(|u_1(t,s)|^2) \leq (1-\lambda_1)^{s-t} |u_1(t)|^2,$$

and

$$\mathsf{E}_u(|u_2(t,s)|^2) \leq (1-\lambda_2)^{s-t} |u_2(t)|^2,$$

and so

$$\begin{aligned} \mathsf{E}_u(|w(t,s)|^2) &= \mathsf{E}_u(|u_1(t,s)|^2 + |u_2(t,s)|^2) \\ &\leq (1-\mu)^{s-t}(|u_1(t)|^2 + |u_2(t)|^2) \\ &\leq (1-\mu)^{s-t} |u(t)|^2. \end{aligned} \tag{1}$$

From the definition of Σ_1 we have

$$\begin{aligned} \mathsf{E}_u(|\Sigma_1(s)|^2) &= \mathsf{E}_u\left(\left|\sum_{t=0}^{s-a} w(t,s)\right|^2\right) \\ &= \mathsf{E}_u\left(\sum_{t=0}^{s-a}\sum_{t'=0}^{s-a} w(t,s)^{\mathsf{T}} w(t',s)\right) \\ &= \sum_{t=0}^{s-a}\sum_{t'=0}^{s-a} \mathsf{E}_u(w(t,s)^{\mathsf{T}} w(t',s)) \\ &\leq \sum_{t=0}^{s-a}\sum_{t'=0}^{s-a} \mathsf{E}_u(|w(t,s)|^2)^{1/2} \mathsf{E}_u(|w(t',s)|^2)^{1/2} \\ &= \left(\sum_{t=0}^{s-a} \mathsf{E}_u(|w(t,s)|^2)^{1/2}\right)^2 \end{aligned}$$

Using (1), this gives

$$\mathsf{E}_u(|\Sigma_1(s)|^2) \leq \left(\sum_{t=0}^{s-a} (1-\mu)^{(s-t)/2} |u(t)|\right)^2 \tag{2}$$

Now we take expectation over the sequence $u(t)$. Since the $|u(t)|$ are independent 0-1 valued variables with mean p, the expectation of the right hand side is easy to estimate:

$$\mathsf{E}\left(\left(\sum_{t=0}^{s-a} (1-\mu)^{(s-t)/2} |u(t)|\right)^2\right)$$

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

$$= \sum_{t=0}^{s-a}\sum_{t'=0}^{s-a}(1-\mu)^{(s-t)/2}(1-\mu)^{(s-t')/2}\mathsf{E}(|u(t)||u(t')|).$$

Here

$$\mathsf{E}(|u(t)||u(t')|) = \begin{cases} p, & \text{if } t = t', \\ p^2, & \text{otherwise.} \end{cases}$$

Thus we can write the sum above as

$$\begin{aligned} & p^2\sum_{t=0}^{s-a}\sum_{t'=0}^{s-a}(1-\mu)^{(s-t)/2}(1-\mu)^{(s-t')/2} \\ & \quad + (p-p^2)\sum_{t=0}^{s-a}(1-\mu)^{s-t} \\ & = p^2(1-\mu)^a\left(\frac{1-(1-\mu)^{(s-a)/2}}{1-(1-\mu)^{1/2}}\right)^2 \\ & \quad + (p-p^2)(1-\mu)^a\frac{1-(1-\mu)^a}{1-(1-\mu)} \\ & \le (1-\mu)^a\left(\frac{4p^2}{\mu^2}+\frac{p-p^2}{\mu}\right) < 5\frac{p}{\mu}(1-\mu)^a. \end{aligned}$$

Using that $\mu \ge 1/\bar{n}_0^4$, the lemma follows. □

We select and fix a set of $2g$ edges as follows. The projection of $\mathbb{R}^E$ to $\mathcal{C}$ is surjective; this implies that there are $2g$ edges $e_1,\dots,e_{2g} \in E$ so that the vectors $\eta(e_1),\dots,\eta(e_{2g})$ form a basis in $\mathcal{C}$. Let $\hat{e}_i \in \mathbb{R}^{E_0}$ be the projection of $\eta(e_i)$ to E_0. It follows from Lemma 6 that $\hat{e}_1,\dots,\hat{e}_{2g}$ form a basis of $\mathcal{L}$. These vectors are rational with denominators at most $n^n f^f < n_0^{n_0}$, so the determinant of this basis is at least $n_0^{-m_0 n_0}$.

Now we are ready to prove the main estimate in this section. Call an experiment *good*, if

(a) at most one excitation occurs in every interval $[(t-1)N, tN]$;

(b) no excitation occurs in any of the intervals $[tN-a, tN]$;

(c) $|z''(t)| \le \delta$ for every t $(1 \le t \le T)$;

(d) for every $i = 1,\dots,2g$, there are at least T' integers t such that exactly one excitation occurs in the interval $[(t-1)N, tN]$, and this is the excitation of the edge e_i.

Lemma 10 *The probability that a run is good is at least* $9/10$.

Proof. The probability that (a) is violated is less than $(Np)^2T \le 1/100$. The probability that (b) is violated is less than $apT \le 1/100$. By Lemma 9, the probability that $|X_1(t)| > \delta/2$ is less than $5n_0^4pe^{-a/n_0^4}/\delta^2$, and hence the probability that there is any t for which $|X_1(Nt)| > \delta/2$ is less than

$$\frac{20pTn_0^4e^{-a/n_0^4}}{\delta^2} \le \frac{1}{100}.$$

If this does not happen, and also (b) holds, then $X_2(Nt) = 0$ for every t, and so

$$\begin{aligned} |z''(t)| &= |y''(Nt) - y''((N-1)t)| \\ &\le |x''(Nt) - x''((N-1)t)| \\ &= |X_1(Nt) - X_1((N-1)t)| \\ &\le |X_1(Nt)| + |X_1((N-1)t)| \le \delta. \end{aligned}$$

Finally, the probability that for a number t we have exactly one excitation is at least $(Np)/2$. The probability that this was edge e_i (an event independent of the timing of the excitation) is at least

$$\frac{1}{m}\frac{Np}{2} \ge \frac{2T'}{T}$$

and so (using that these events are independent for different intervals), the probability that it happens for fewer than than T' values of t is less than $1/80g_0$. The probability that this happens for at least one i is less than $1/40$. □

We note that if the run is good, then for every t, we have

$$\begin{aligned} |z'(t)| &= |y'(Nt) - y'((N-1)t)| \\ &\le |x'(Nt) - x'((N-1)t)| \le 1, \end{aligned} \qquad (3)$$

since $|x'(Nt) - x'((N-1)t)|$ is either 0 or the projection of the increment of a single excitation step to the subspace $\mathcal{C}$.

4.2 Completing the proof

Now we are ready to complete the proof. Let k be the value returned by the algorithm. To simplify notation, put $z_i = z(t_i)$, $z_i' = z'(t_i)$ and $z_i'' = z''(t_i)$. It follows from the choice of the integers t_i that the vectors $z_1, z_2, \dots z_k$ are linearly independent, and so $\dim(\mathcal{L}(i)) = i$. Furthermore, by the selection of

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

these vectors, the Gram-Schmidt orthogonalization $z_1^* = z_1, z_2^*, \ldots, z_k^*$ consists of vectors of length at least ε.

Claim 1. If the run is good, then $k \leq 2g$.

It suffices to prove that the vectors $z_1', \ldots, z_k'$ are linearly independent, since they belong to a $(2g)$-dimensional space. We may assume that $k \leq 2g+1$.

Assume that there is a linear relation $\sum_i \alpha_i z_i' = 0$ where not all the α_i are 0. We can write this as $\sum_i \alpha_i z_i = \sum_i \alpha_i z_i''$. Since the z_i are linearly independent, the left hand side is non-zero, and so we may assume it has norm 1:

$$\left|\sum_{i=1}^{k} \alpha_i z_i\right| = \tag{4}$$

We claim that

$$|\alpha_i| \leq \frac{1}{\varepsilon}\left(1+\frac{1}{\varepsilon}\right)^{k-i}. \tag{5}$$

Indeed, (4) implies that for any $1 \leq j \leq k$,

$$\left|\sum_{i=1}^{k} \alpha_i z_i^{\mathsf{T}} z_j^*\right| \leq |z_j^*|.$$

Since $z_i^{\mathsf{T}} z_j^* = 0$ if $i < j$ and $z_j^{\mathsf{T}} z_j^* = |z_j^*|^2$, it follows that

$$|\alpha_j||z_j^*|^2 \leq |z_j^*| - \sum_{i=j}^{k} |\alpha_i z_i^{\mathsf{T}} z_j^*| \leq |z_j^*| + \sum_{i=j}^{k} |\alpha_i||z_j^*|.$$

Dividing by $|z_j^*|^2$ and using that $|z_i| \leq 1$, we get that

$$|\alpha_j| \leq \frac{1}{|z_j^*|}\sum_{i=j}^{k} |\alpha_i| \leq \frac{1}{\varepsilon}\sum_{i=j}^{k} |\alpha_i|.$$

Hence (5) follows by induction on $k-i$.

Now using (5) and $|z_i''| \leq \delta$, we get

$$\left|\sum_i \alpha_i z_i''\right| \leq \delta \sum_i |\alpha_i| \leq \delta\left(1+\frac{1}{\varepsilon}\right)^k < \delta\varepsilon^k e^{k\varepsilon} < 1,$$

a contradiction. This proves Claim 1.

Claim 2. The probability that the run is good but $k < 2g$ is less than $1/10$.

Suppose that $k < 2g$. Since the determinant of the vectors $\mu(e_1)\ldots,\mu(e_{2g})$ is at least $n_0^{-m_0 n_0} = 10\varepsilon$, it follows by an easy geometric argument that at least one of these vectors, say $\mu(e_1)$, is at a distance at least 2ε from $\mathcal{L}(k)$. We assumed that the run was good, and hence there were at least T' integers t for which exactly one excitation occurred in the interval $[(t-1)N+1, tN]$, and this was the excitation of edge e_1. For such a t, we have $z'(t) = \mu(e_1)$, and so

$$d(z(t), \mathcal{L}(k)) \geq d(z'(t), \mathcal{L}(k)) - |z''(t)| \geq \varepsilon.$$

Thus we were not supposed to stop yet, a contradiction.

It follows that the probability that $k = 2g$ is at least $4/5$.

5 Variations

We can make some cosmetic changes to the setup as given above. One objection may be that the noisy circulator, as defined, is not truly local, since (say) in operation (a) we have to select a node uniformly from all nodes. The standard way of fixing this is to attach an "alarm clock" to each node, edge, and face, which wakes them up at random times according to a Poisson process (the edge-clock is much slower than the other two).

Another objection is that the noisy circulator as constructed above is not stationary: the total mass grows to infinity. An easy fix is to give a second, even slower clock to each edge: when this rings, they reset their value to 0. Another possible fix comes from the observation that the excitations don't necessarily have to be constants, the proof works just as well for random and symmetric excitations. So modifying the excitation step to reset the value of any edge to 1 with very small probability (rather than to add 1) provides a stationary version (but the analysis becomes more complicated).

Further variants, improvements and generalizations of the above system are of interest:

- We had to put a rather tight upper bound on the excitation rate p (its only virtue is that it is polynomial in the data). Can one recover the genus even if the rate at which excitations

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

take place is faster (ideally, independent of the number of nodes)?

- Suppose that we only allow two values (or any other given discrete set of values) on any edge. Can we still recover the genus?
- Can one extend the technique to recover a non-orientable surface by observing a random process on an embedded graph locally?

For background regarding graphs on surfaces and Riemann surfaces see for instance [13] [4], for background regarding algorithmic applications of random processes see [8].

Instead of a passive observer, we can also consider an active observer, who can impose an excitation on any of the edges in E_0. In this case, we assume that spontaneous excitations don't occur. The analysis above goes through, and it becomes even easier: we know exactly when the last excitation occurred, and how long we have to wait until $x(t)$ settles down. The only point where we need to add something to our analysis above is before the definition of a "good run", where we need to argue that there are $2g$ edges in E_0 (rather than in E) so that the vectors $\eta(e_1),\ldots,\eta(e_{2g})$ form a basis in $\mathcal{C}$.

Suppose that this is not so. Then there is a nonzero vector $w \in \mathcal{C}$ so that w is orthogonal to every vector $\eta(e)$, $e \in E_0$. Write $\eta(e) = T\chi(e)$, where T is the projection map onto $\mathcal{C}$. But then by the symmetry of T we have

$$0 = \eta(e)\cdot w = (T\chi(e))\cdot w = \chi(e)\cdot(Tw) = \chi(e)\cdot w,$$

and hence w is 0 on every edge in E_0. By Theorem 6, this is impossible.

6 Other models

The notion of global information from local observation can inspire many questions in different directions. We briefly present some related examples.

Example 1 [Checking for a perfect matching.] Putting together ideas from [10] and [5], we can formulate the following randomized distributed algorithm to find a maximum matching in a bipartite graph. Every edge carries a nonnegative weight, which at the start is 1. At each step, we pick a random node v, which does the following: (a) if at least one edge in δv has positive weight, it rescales the edges in δv, so that their sum becomes 1; (b) with some small probability p, if there is more than one positive weight on the edges in δv, then it reduces the smallest one to 0. In polynomial expected time, we end up with the edges of a maximum matching having weight 1, all the other edges having weight 0. (Details of this result, which is joint work with Mike Saks, will be published elsewhere.)

We can turn this algorithm to an example of our main theme by allowing negative weights, and adding a third rule: (c) if v sees only 0 weights, or at least one negative weight, then it reduces all the weights in δv to -1. Then it suffices to observe any given edge: if the graph has no perfect matching, then this edge turns negative within polynomial time with large probability; else, it stays nonnegative with large probability. Thus we can determine the existence of a perfect matching from local observation.

Example 2 [Observing returns of a random walk.] Now we turn our approach around: we fix the local random process, and want to determine which properties of the graph can be recovered from local observation. Let G be a finite connected graph, and consider a simple random walk on G. Fix a vertex v in G. You are given the sequence of times when the simple random walk visits v; what information can you learn about the graph G? (A spelunker has an accident in the depth of a cave; without a light and unable to move, all he can sense is a bat passing by every now and then on its erratic flight through the cave. What can he learn about the shape of the cave?)

An example of Gábor Tardos shows that the graph cannot be determined, even if we know that it is a tree, and we can observe the whole infinite sequence of return times. In fact, even the spectrum of the transition matrix of the random walk on the graph is not determined: in Tardos's example, the two trees have the same eigenvalues, but with different multiplicities. One positive result: if we know that the graph is has a node-transitive automorphism group, then the eigenvalues can be determined (but not necessarily their multiplici-

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

ties). We don't know if the eigenvalues (without multiplicities) can be determined for general graphs. Another, perhaps more important twist on the problem: suppose that we know that the eigenvalues can be determined (e.g. for a node-transitive graph); to what precision can they be determined if we only observe the process for a polynomial time?

Example 3 [Obtaining the size of the road system by measuring the volume of local traffic]. Let G be a finite connected irreducible regular graph. From each vertex start an independent simple random walk. Fix a vertex $v \in G$. For each time t you can observe the number of walkers occupying v. How much time is needed in order to, almost surely, know the size of G? Assume you are given an *a priori* bound n_0 on $n = |V(G)|$. Then one can get a (poor) polynomial upper bound on the time needed along the following lines. The mixing time of G is smaller than n^3, so if we observe the load on v only at times of the form kn_0^3, we get almost independent samples from a distribution which is exponentially close to Binomial$(n, 1/n)$, the stationary distribution for the number of walkers at v. For Binomial$(n, 1/n)$, about n^4 samples are needed to recover n. So after about n_0^7 steps n can be recovered. (Probably n_0^4 is the time needed to recover n.)

Further examples of interesting local random processes (for which we know very little about information obtainable from local observation) include the heat-bath chain on graph colorings, various avalanche (chip-firing) models, or the token permutation chain [1].

Acknowledgement. We are grateful to Dimitris Achlioptas, Mike Freedman, Eran Makover, Mike Saks, Oded Schramm and Kevin Walker for stimulating discussions on this topic.

References

[1] D. Aldous, My favorite 6 open problems in mathematical probability. Preprint, (2001).

[2] B. Bollobás, Volume estimates and rapid mixing. *Flavors of geometry*, 151–182, Math. Sci. Res. Inst. Publ., **31**, Cambridge Univ. Press, Cambridge, 1997.

[3] R.L. Brooks, C.A.B. Smith, A.H. Stone, W.T. Tutte, The dissection of rectangles into squares. Duke Math. J. 7, (1940). 312–340. (Reviewer: P. Scherk) 48.0X

[4] P. Buser, Geometry and spectra of compact Riemann surfaces. Progress in Mathematics, 106. Birkheuser Boston, Inc., Boston, MA, xiv+454 pp, 1992.

[5] J. Csima and L. Lovász: A matching algorithm for regular bipartite graphs, *Discrete Appl. Math.* **35** (1992), 197–203.

[6] R.J. Duffin, Basic properties of discrete analytic functions, *Duke Math. J.* **23** (1956), 335–363.

[7] J. Ferrand, Fonctions préharmoniques et fonctions préholomorphes. *Bull. Sci. Math.* **68**, (1944). 152–180.

[8] O. Häggström: Finite Markov chains and algorithmic applications. Preprint, (2002).

[9] H. Kesten: Distinguishing and reconstructing sceneries from observations along random walk paths, *Microsurveys in Discrete Probability* DIMACS Ser. Discrete Math. Theoret. Comput. Sci. **41**, Amer. Math. Soc., Providence, RI (1998), 75–83.

[10] N. Linial, A. Samorodnitsky and A. Wigderson: A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents, *Combinatorica* **20** (2000), 545–568.

[11] L. Lovász and P. Winkler, Mixing times, *Microsurveys in Discrete Probability* DIMACS Ser. Discrete Math. Theoret. Comput. Sci. **41**, Amer. Math. Soc., Providence, RI (1998), 85–133.

[12] C. Mercat, Discrete Riemann surfaces and the Ising model. *Comm. Math. Phys.* **218** (2001), no. 1, 177–216.

[13] B. Mohar and C. Thomassen, Graphs on surfaces. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, xii+291 pp, 2001.

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 **IEEE**