Describing lengths of walks in directed graphs

Anthony Widjaja Lin

Oxford University Department of Computer Science

Introduction

Problem: Given a (directed) graph $G = (V, E)$ and two nodes $s, t \in V$, describe the set $L(G, s, t)$ of lengths of walks from s to t in a "nice" way.

- ✔ Self-loops are allowed in our graphs.
- ✔ "Nice" means can be easily manipulated algorithmically
  - ✘ So, infinite sums are less "nice"
- ✔ "Nice" means easy to understand.

Introduction

Problem: Given a (directed) graph $G = (V, E)$ and two nodes $s, t \in V$, describe the set $L(G, s, t)$ of lengths of walks from s to t in a "nice" way.

- ✔ Describe $L(G, s, t)$ as a union of a small number of arithmetic progressions (AP) with small offsets and periods.
  - ✘ AP is any set $a + b\mathbb{N} := \{a + bk : k \in \mathbb{N}\}$, for some offset $a \in \mathbb{N}$ and period $b \in \mathbb{N}$.
- ✔ The result has interesting applications in computer science. (trust me)

Chrobak-Martinez Theorem

Theorem (Chrobak-Martinez): Given a graph $G = (V, E)$ with nodes $s, t \in V$, the set $L(G, s, t)$ can be described as a union of $O(|V|^2)$ many arithmetic progressions with offsets at most $O(|V|^2)$ and periods at most $O(|V|)$. Moreover, this union can be computed in polynomial time.

- ✔ Describable as a union of exponentially many APs is well-known in automata theory (using "determinization": compute large "equivalent" lasso-shaped graph).
- ✔ CM Theorem gives exponentially better upper bounds.
- ✔ Matching (asymptotic) lower bounds are known.

Chrobak-Martinez Theorem

Theorem (Chrobak-Martinez): Given a graph $G = (V, E)$ with nodes $s, t \in V$, the set $L(G, s, t)$ can be described as a union of $O(|V|^2)$ many arithmetic progressions with offsets at most $O(|V|^2)$ and periods at most $O(|V|)$. Moreover, this union can be computed in polynomial time.

- ✔ A subtle bug in the original proof was recently found and fixed.
- ✔ This talk describes a corrected proof. (A restriction of the proof of a more general case in my LICS'10 paper)

Outline of proof

- ✔ Little lemma from number theory about solutions of Diophantine equations.
- ✔ Graph analysis to derive existence of APs descriptions.
- ✔ Further analysis to derive polynomial-time algorithms.

A few words on notations and terminologies

- ✔ Given two sets $S_1, S_2 \subseteq \mathbb{N}$, define standard arithmetic operations on such sets, e.g.,:
  - ✘ $S_1 + S_2 := \{s_1 + s_2 : s_1, s_2 \in \mathbb{N}\}$.
  - ✘ $S_1 \cdot S_2 := \{s_1.s_2 : s_1, s_2 \in \mathbb{N}\}$.
- ✔ If S_1 is a singleton $\{s_1\}$, write $s_1 + S_2$ (resp. $s_1 \cdot S_2$) for $\{s_1\} + S_2$ (resp. $\{s_1\} \cdot S_2$).
- ✔ This notation is consistent with earlier definition of an arithmetic progression $a + b\mathbb{N}$.
- ✔ A closed walk in G is short if it is of length $\leq |V|$.

A little lemma from number theory

Lemma (Erdos-Graham): For all $v_1, \ldots, v_m \in \mathbb{N}$, each set

$$v_1\mathbb{N} + \cdots + v_m\mathbb{N}$$

with $n := \max(v_1, \ldots, v_m)$ equals $S \cup a + b\mathbb{N}$, where

- ✔ each $c \in S$ is $\leq n^2$,
- ✔ a is the least integer bigger than n^2 that is a multiple of $b := \gcd(v_1, \ldots, v_m)$

If $\gcd(v_1, \ldots, v_m) = 1$, then eventually all numbers are in $v_1\mathbb{N} + \cdots + v_m\mathbb{N}$.

A little lemma from number theory

Lemma (Erdos-Graham): For all $v_1, \ldots, v_m \in \mathbb{N}$, each set

$$v_1\mathbb{N} + \cdots + v_m\mathbb{N}$$

with $n := \max(v_1, \ldots, v_m)$ equals $S \cup a + b\mathbb{N}$, where

- ✔ each $c \in S$ is $\leq n^2$,
- ✔ a is the least integer bigger than n^2 that is a multiple of $b := \gcd(v_1, \ldots, v_m)$

Can be computed in polynomial time (numbers given in unary).

A little lemma: example

Suppose that McDonalds sells their Chicken McNuggets in 3s and in 7s. What is the largest number of Chicken McNuggets that cannot be purchased?

A little lemma: example

$3\mathbb{N} + 7\mathbb{N}$ equals

$$0 + 3\mathbb{N} \bigcup 7 + 3\mathbb{N} \bigcup 14 + 3\mathbb{N}$$

How so?

A little lemma: example

$3\mathbb{N} + 7\mathbb{N}$ equals

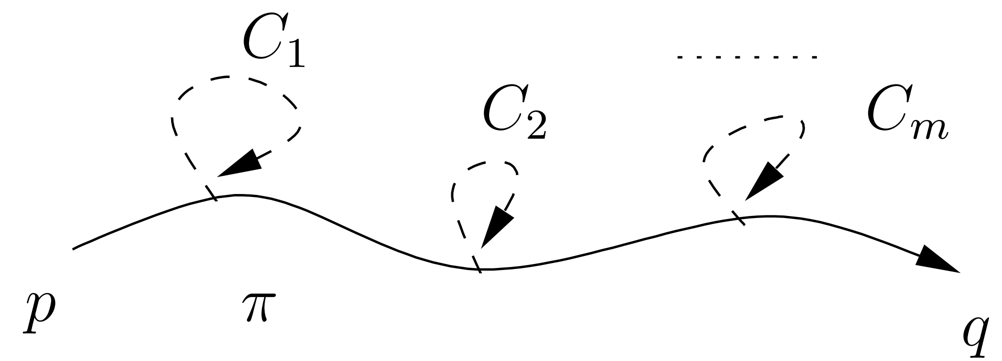
$$0 + 3\mathbb{N} \bigcup 7 + 3\mathbb{N} \bigcup 14 + 3\mathbb{N}$$

How so?

Therefore, $3\mathbb{N} + 7\mathbb{N} = \{0, 3, 6, 7, 9, 10\} \cup 12 + 1\mathbb{N}$. So, the answer is smallest $2 \pmod 3$ representative less than 14, which is 11.

Describing $L(G, s, t)$ as a union of APs

Take a walk π and all short closed walks $C_1, \ldots, C_m$ visiting π.



The walk type T_π of π is the linear set $|\pi| + \sum_{i=1}^{m} |C_i|\mathbb{N}$.

Removing redundant terms, T_π has "small" size

Describing $L(G,s,t)$ as a union of APs

Lemma: $L(G,s,t) = \bigcup_{\pi} T_{\pi}$, where π ranges over walks from s to t of length at most $(|V|-1)^2$.

Applying Erdos-Graham on each T_{π}, we get 1 part of CM:

Corollary: $L(G,s,t)$ coincides with the union $I \cup K$ s.t.

- ✔ I contains no numbers exceeding $|V|^2 + (|V|-1)^2$,
- ✔ K is a union of arithmetic progressions of the form $a + b\mathbb{N}$ with $|V|^2 + (|V|-1)^2 < a \leq |V|^2 + (|V|-1)^2 + |V|$ and $0 < b \leq |V|$.

Describing $L(G,s,t)$ as a union of APs

To prove this characterization lemma, we need the following simple fact.

Fact: Given a walk π in G from p to q, there exists a (simple) path π' from p to q and finitely many cycles $C_1, \ldots, C_h$ (possibly with duplicates) such that

$$|\pi| = |\pi'| + \sum_{i=1}^{h} |C_i|$$

Note: the proof is similar to the existence of simple paths between two connected vertices.

Describing $L(G, s, t)$ as a union of APs

Let us prove $L(G, s, t) \subseteq \bigcup_\pi T_\pi$.

Take $a \in L(G, s, t)$ and a witnessing walk π from s to t.

Describing $L(G, s, t)$ as a union of APs

Let us prove $L(G, s, t) \subseteq \bigcup_\pi T_\pi$.

Take $a \in L(G, s, t)$ and a witnessing walk π from s to t.

Q: Why not just use Fact obtaining the simple path π' + cycles, and show that $a \in T_{\pi'}$?

Describing $L(G, s, t)$ as a union of APs

Let us prove $L(G, s, t) \subseteq \bigcup_{\pi} T_{\pi}$.

Take $a \in L(G, s, t)$ and a witnessing walk π from s to t.

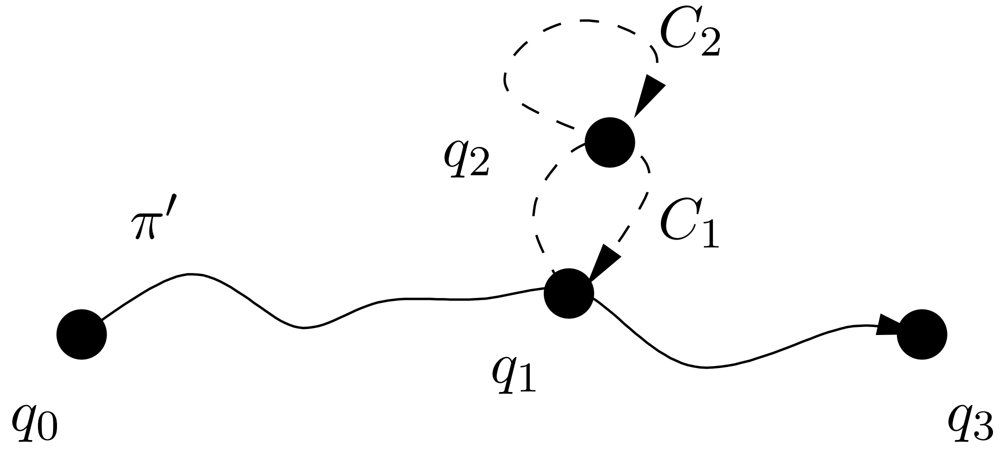
Q: Why not just use Fact obtaining the simple path $\pi' +$ cycles, and show that $a \in T_{\pi'}$?
A: Cannot

Describing $L(G, s, t)$ as a union of APs

Let us prove $L(G, s, t) \subseteq \bigcup_{\pi} T_{\pi}$.

Take $a \in L(G, s, t)$ and a witnessing walk π from s to t.

Q: Why not just use Fact obtaining the simple path $\pi' +$ cycles, and show that $a \in T_{\pi'}$?
A: Cannot



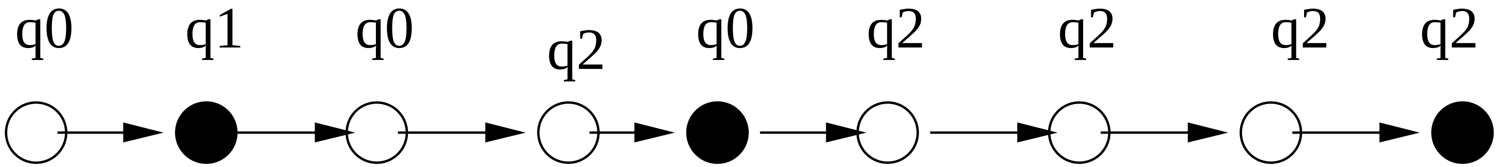
Describing $L(G, s, t)$ as a union of APs

Let us prove $L(G, s, t) \subseteq \bigcup_\pi T_\pi$.

Take $a \in L(G, s, t)$ and a witnessing walk π from s to t.

Q: Why not just use Fact obtaining the simple path $\pi' +$ cycles, and show that $a \in T_{\pi'}$?
A: Cannot

Soln: For each node p in π, pick the last position in π at which p is visited + apply Fact on each segment.



The resulting walk is of length at most $(|V| - 1)^2$.

Describing $L(G,s,t)$ as a union of APs

Conversely, that $L(G,s,t) \supseteq \bigcup_{\pi} T_{\pi}$ is obvious from definition of T_{π}.
Any $a \in T_{\pi}$ describes the length of a walk from s to t in G.

Computational complexity

Summary so far:

Corollary: $L(G, s, t)$ coincides with the union $I \cup K$ s.t.

- ✔ I contains no numbers exceeding $|V|^2 + (|V| - 1)^2$,
- ✔ K is a union of APs of the form $a + b\mathbb{N}$ with $|V|^2 + (|V| - 1)^2 < b \leq |V|^2 + (|V| - 1)^2 + |V|$ and $0 < b \leq |V|$.

A naive implementation of the previous construction takes exponential time.

Can do better using dynamic programming!

Computational complexity

- ✔ Characteristic Matrices $\mathbb{M} = \{M_i\}_{i=0}^{n-1}$ of existence of walks of length $i = 1, \ldots, n-1$ can be computed in cubic time. (by matrix multiplication)

- ✔ So, for each state q, the set Γ_q of lengths of closed walks meeting with q up to length n can be read off directly from $\mathbb{M}$. (diagonal entries)

- ✔ For each $v \in V$, let $T_{i,v} := \bigcup_\pi T_\pi$, where π ranges over walks in G of length i from s to v.

- ✔ So, $L(G, s, t) = \bigcup_{i=0}^{(|V|-1)^2} T_{i,t}$.

Computational complexity

We can compute each $\{T_{i,v}\}_{v\in V}$ recursively on i using:

Lemma: $T_{0,s} = \{0\}$ and $T_{0,v} = \emptyset$ for $v \in V \setminus \{s\}$.
Whenever $i > 0$ and $v \in V$, we have:

$$T_{i,v} = \bigcup_{u\in V} \left(T_{i-1,u} + \sum_{(u,v)\in E, k\in\Gamma_v} (1 + k\mathbb{N}) \right) .$$

Each application of this lemma is to be combined with Erdos-Graham (and remove redundant linear sets) to keep the size of $T_{i,v}$ small.

This takes polynomial time.

Algorithm in action on a wee example

See the white/black board.

An application

Let IP_k be the set of integer programs with variables $x_1, \ldots, x_k, y$. For every fixed k, checking whether $P \in \mathrm{IP}_k$ has an integer solution is solvable in polynomial time (Lenstra 81).

Problem: Given $P \in \mathrm{IP}_k$, a graph $G = (V, E)$, and nodes $s, t \in V$, decide the existence of a walk from s to t in G of length k such that P is solvable with $y := k$.

This problem is solvable in polynomial time by first computing a union of m APs for $L(G, s, t)$, and solve m integer programs in IP_{k+1} (each obtained by "appending" each AP to P).

Other computer science applications

- ✔ Polynomial-time algorithms for checking walks with more complex constraints (motivation from databases)
- ✔ An optimal complexity for model checking EF logic over one-counter machines. (infinite-state model checking)

Bibliographic notes

- ✔ The problem we considered is an instance of a more general problem in automata theory.
- ✔ An automaton is an edge-labeled graph $G = (V, \{E_a\}_{a\in\Sigma})$, where $\Sigma := \{a_1, \ldots, a_k\}$ is a finite set (called alphabet).
- ✔ Given $s, t \in V$, let $L(G, s, t)$ denote the set of tuples $(m_1, \ldots, m_k)$ such that there is a path from s to t in which each $a_i \in \Sigma$ occurs precisely m_i times.
- ✔ Can we describe $L(G, s, t)$ in a "nice" way?

Bibliographic notes

- ✔ Can we describe $L(G, s, t)$ in a "nice" way? Yes
- ✔ Use generalizations of union of arithmetic progressions (called semilinear sets)
- ✔ Chrobak-Martinez Theorem is replaced by Caratheodory Theorem from convex geometry.
- ✔ Use bounds on solutions of integer programs (Papadimitriou'83).
- ✔ Lots of applications in computer science and operation research.
- ✔ See my LICS'10 paper for this.

Thank you

Thank you in every language.