Notes for the Lectures on May 8th and May 11th

Kurt Mehlhorn

May 11, 2009

This is an extended version of pages 29, 30 and 31 of [KLM+09]

Require: G is a connected graph with n_0 nodes and m_0 edges;
$G_c = G$;
declare all nodes unlabeled;
initialize the basis to the empty set;
$\{$let $N_c = m_c - (n_c - 1)\}$
while G_c is not a tree **do**
$\{G_c$ is connected and not a tree$\}$
while G_c has a node of degree 1 **do**
remove it and the incident edge; declare the removed edge a tree edge;
$\{m_c$ and n_c are decreased by one and N_c does not change$\}$
end while
$\{G_c$ is connected, not a tree and every node has degree at least two$\}$
if every node of G_c has degree two, i.e., G_c is a circuit **then**
add this circuit to the basis, declare one of its edges non-tree and delete it from G_c;
$\{m_c$ and N_c went down by one; G_c is now a tree$\}$
else
$\{G_c$ is connected, not a tree, and there is a node of degree at least three$\}$
construct an auxiliary graph; its nodes correspond to the nodes in G_c of degree at least three and its edges correspond to the maximal paths in G_c with all internal edges having degree two;
let C_a be a circuit in the auxiliary graph consisting of at most $1+2\log n_0$ auxiliary edges;
add the underlying circuit in G_c to the basis, and delete all edges comprising the heaviest auxiliary edge on this circuit from G_c; declare one of these edges non-tree and all others tree;
$\{m_c$ and N_c went down by one and the weight of the circuit added to the basis is at most $(1+2\log n_0)$ times the weight of the edges deleted$\}$
end if
end while $\{G_c$ is a tree and hence $N_c = 0\}$
declare all edges of G_c tree edges and delete them from the graph;

Lemma 1 *The total weight of the circuits is at most $(1+2\log n_0)W$ where $W = \sum_e w(e)$ is the total weight of all edges.*

Proof: For every circuit added to the basis, its weight is at most $(1+2\log n_0)$ times the weight of the edges deleted. Observe that this is also true for the last circuit removed (one of its edges is removed in the while-loop and the others are deleted after the while-loop). Thus the total weight of all circuits added to the basis is at most $(1+2\log n_0)W$. ∎

Lemma 2 *The number of circuits constructed and the number of edges declared non-tree is* $m_0-(n_0-1)$.

Proof: Consider the quantity $N_c = m_c-(n_c-1)$, where n_c and m_c are the number of nodes and edges of the current graph, respectively. N_c starts at $m_0-(n_0-1)$ and ends at 0. Removal of a vertex of degree one, does not change N_c, addition of a circuit to the basis decreases it by one. Thus we add exactly $m_0-(n_0-1)$ circuits to the basis. For each circuit constructed, we declare one edge non-tree. ∎

Lemma 3 *Let* Γ *be the cycle matrix corresponding to the basis constructed where we order the circuits in their order of construction and the non-tree edges in the order in which they are declared non-tree. Then the square submatrix* Γ' *of* Γ *selected by the non-tree edges is a lower triangular matrix. Each diagonal entry is either* $+1$ *or* -1. *The determinant of* Γ' *is* ± 1.

Proof: Let $C_1, \ldots, C_N$ be the circuits in the order in which they are constructed and $e_1, \ldots, e_N$ the edges declared non-tree in the order in which they are declared non-tree. Then C_i uses e_i and hence each diagonal entry is either $+1$ or -1. Also, e_i is deleted after the construction of C_i and hence $C_j(e_i) = 0$ for $j > i$. Thus the elements above the diagonal are zero. ∎

Lemma 4 *The edges designated as tree edges form a spanning tree.*

Proof: Observe first that we designate $m_0-(n_0-1)$ edges as non-tree and hence n_0-1 edges as tree. The edges designated non-tree select a non-singular submatrix of Γ. Hence the edges designated tree form a spanning tree. ∎

Theorem 1 *The algorithm constructs an integral basis of weight* $O(W\log n)$.

Proof: We have already shown the weight bound.

Let C be any cycle. We need to show that C is a integer linear combination of our circuits, i.e., $C = \Gamma x_C$ for an integral vector x_C. Let C' and Γ' be the restrictions to the non-tree edges. Then $C' = \Gamma' x_C$. Cramer's rule implies that the entries of x_C are rational numbers whose entries have denominator $\det \Gamma'$. Thus x_C is integral. ∎

How good is the bound of Theorem 1? Can we do better? We approach this question from several directions.

1. In the case of uniform weights, i.e., $w(e) = 1$ for all e, we can improve upon the bound for graphs with a non-linear number of edges. We will show that any graph has an integral basis of total cardinality $O(m(\log n)/\max(1, \log(m/n)))$.

2. We show that the bound in item 1 is optimal.

3. We show (exercise sheet 2) that a complete graph has a basis of weight $O(W)$.

4. We pose an open problem.

Theorem 2 *Any graph has an integral basis of total cardinality* $O(m\frac{\log n}{\max(1,\log(m/n))})$.

Proof: We need the following lemma. A beautiful proof can be found in [AHL02]. In exercise sheet 2, we prove the result for regular graphs of degree $d = m^{1/k}$.

Lemma 5 *Let* $k \geq 2$. *Any graph with* $m \geq n^{1+1/k}$ *edges contains a circuit of length* $O(k)$.

If $m \leq 2n$, Theorem 1 does the job. It yields a basis of length $O(m \log n)$. So assume that $m > 2n$. Let $k = 2\log n/\log(m/n)$. We proceed in two phases.

- As long as $m \geq n^{1+1/k}$, we find a circuit of length $O(k)$, add it to the basis and delete one of its edges from the graph. The total length of the circuits added in phase I is $O(mk)$.

- If $m \leq n^{1+1/k}$, we apply Theorem 1 and obtain a basis of total length $O(n^{1+1/k}\log n)$ for the remaining graph.

The total length of the basis is $O(km + n^{1+1/k}\log n)$. Finally,

$$\frac{n^{1+1/k}\log n}{km} = \frac{n 2^{\log n \frac{\log(m/n)}{2\log n}} \log n \log(m/n)}{m 2 \log n} = \frac{n 2^{\frac{\log(m/n)}{1/2}} \log(m/n)}{2m}$$

$$= \frac{n\sqrt{\frac{m}{n}}\log(m/n)}{2m} = \frac{m\sqrt{\frac{n}{m}}\log(m/n)}{2m} = O(\frac{\log(m/n)}{\sqrt{m/n}}) = O(1) .$$

∎

Discussion: why this choice of k? give upper bounds for special values of m, say $m = \Theta(n)$, $\Theta(m \log n)$, and $\Theta(n^{1+1/k})$.

Exercise 1 *Consider arbitrary non-negative edge weights? Why doesn't the proof above show that any graph has a basis of weight* $O(W\frac{\log n}{\log(m/n)})$?

We next prove a lower bound.

Theorem 3 *Let* $k \geq 2$. *For sufficiently large* n, *there is a graph with* $\Theta(n^{1+1/(2k)})$ *edges (the claim is actually true with* $2k$ *replaced by* k*) and no circuit of length shorter than* k. *In such a graph any cycle basis has total length* $\Omega(m\frac{\log n}{\log(m/n)})$.

Proof: Assume first such a graph exists. In this graph any cycle basis has length at least $(m-n+1)k = \Omega(mk)$. Also, $m = \Theta(nn^{1/(2k)})$ or $\log(m/n) = \Theta((1/2k)\log n)$ or $2k = \Theta(\frac{\log n}{\log(m/n)})$.

We next show the existence of the graph. We sketch a proof by Erdös from 1957; it is one of the first examples of the so-called probabilistic method [ASE92]. We will NOT construct a graph with the claimed properties, we will only show the existence.

Let $\varphi = 1/(2k)$ and consider a random graph $G(n,p)$ with $p = n^{\varphi-1}$. In such a graph, each of the $n(n-1)/2$ potential edges is present with probability p. For a G in $G(n,p)$,

- the expected number of edges is $pn(n-1)/2 \approx 1/2n^{1+\varphi}$.
- for each node the expected degree is $p(n-1) \approx n^{\varphi} = n^{1/(2k)}$.

For almost all graphs in $G(n,p)$, all but a fraction $o(1/n)$, the number of edges is at least $1/4n^{1+\varphi}$ and the degree of every node is at most $2n^{\varphi}$.

Let X be the number of circuits of length less than k. Then

$$\mathrm{E}[X] = \sum_{3 \le i < k} \frac{(n)_i}{2i} p^i \le \sum_{i<k} (np)^i = \frac{(np)^k - 1}{np - 1} \le (np)^k = n^{\varphi} = \sqrt{n}\ ,$$

where the last inequality uses the fact that $np = n^{\varphi} \ge 2$ for n large enough.

Thus there is a graph in $G(n,p)$ satisfying the two items above and having only $\sqrt{n}$ circuits of length less than k. We remove one node from each such circuit and obtain a graph G' with

- n' nodes, where $n' \le n$, and
- m' edges, where $m' \ge m - \sqrt{n}2n^{1/(2k)} \ge (1/4)n^{1+\varphi} - 2n^{1/2+\varphi} \ge 1/8n^{1+\varphi}$.
- no circuit of length less than k.

∎

Problem 1 *Do sufficiently dense graphs always have a cycle basis of weight $o(W \log n)$? Observe that complete graphs have cycle basis of weight $O(W)$ (exercise sheet 2).*

KM conjectures that the answer is yes.

References

[AHL02] N. Alon, S. Hoory, and N. Linial. The Moore bound for irregular graphs. *Graphs and Combinatorics*, 18:53–57, 2002.

[ASE92] N. Alon, J.H. Spencer, and P. Erdös. *The Probabilistic Method*. John Wiley & Sons, 1992.

[KLM+09] T. Kavitha, Ch. Liebchen, K. Mehlhorn, D. Michail, R. Rizzi, T. Ueckerdt, and K. Zweig. Cycle Bases in Graphs: Characterization, Algorithms, Complexity, and Applications. 78 pages, submitted for publication, available at `www.mpi-inf.mpg.de/~mehlhorn/ftp/SurveyCycleBases.pdf`, March 2009.