Remarks on Matchings

Kurt Mehlhorn

MPI für Informatik

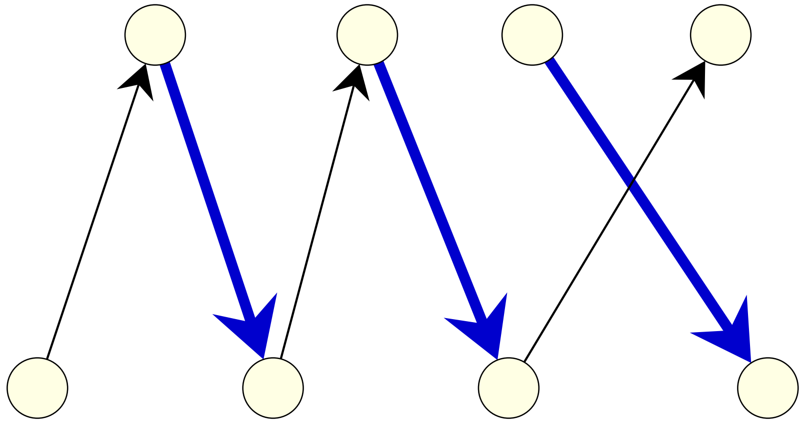
Saarbrücken

Germany

all papers are available from my home page

Matchings

- $G = (A \dot{\cup} B, E)$, bipartite graph
- matching M = subset of edges no two of which share an endpoint



- matching problems are abundant: males and females, persons and jobs, families and houses, medical students and hospitals, students and lab sessions, professors and offices, clients and servers
- goal: find best matching (assignment) in some sense

Optimization Criteria I

- maximum cardinality matching

$$\text{maximize } |M|$$

- maximum weight matching
  - each edge e has a weight (profit, utility) $w(e)$
  - maximize the total weigth of the matching

$$\text{maximize } \sum_{e \in M} w(e)$$

Optimization Criteria II

- Economics, Social Sciences
  - nodes in A (and B) rank their incident edges: I prefer x over y or I am indifferent between x and y
  - ranking = linear order without or with ties
  - one side ranks
    - professors rank offices, persons rank jobs,...
  - both sides rank
    - females rank males and males rank females
    - medical interns rank hospitals and hospitals rank medical interns
  - students rank potential roommates (general graph)
  - hospitals have capacity larger than one
  - rich source of problems with practical relevance and theoretical appeal
  - many sensible optimization criteria

Structure of Talk

- Part I: only one side ranks
- the theme will be different notions of optimality

- Part II: both sides rank
- the theme will be stability

- Part III: cardinality matching
- the theme will be average case behavior

- I got interested in 3) because I presented a paper by Motwani in class, R. Irving and D. Abraham introduced me to 1), and a Google search for strange time bounds (here $O(m^2)$) got me into 2)

One Side Ranks

- the nodes in A assign integer ranks to their indicent edges
- $E = E_1 \dot{\cup} E_2 \ldots \dot{\cup} E_r$
- $E_i =$ edges of rank i
- no ties: E_i contains at most one edge incident to any $a \in A$.
- What are sensible notions of optimality?
  - pareto-optimality
  - popularity
  - rank maximality

Pareto-Optimal Matchings

- M is Pareto-optimal if there is no N in which no node is worse off and at least one node is better off

- are maximal, but may have different cardinalities

- characterization of Pareto-optimal matchings
  - maximal: if a is unmatched in M, all potential partners are matched
  - trade-in-free: if a is matched in M, all better partners are also matched
  - coalition-free: no cycle of exchanges, in which no one looses and one wins

- minimum cardinality Pareto-optimal is NP-complete
  reduction from minimum maximal matching

- maximum cardinality in time $O(\sqrt{n}m)$
  compute maximum matching and convert into POM of same card

ISAAC 04, joint work with D. Abraham, K. Cechlárová, D. Manlove

Popular Matchings

- M is more popular than N if the number of nodes preferring M over N is larger than the number of nodes preferring N over M
- popular matching = no matching which is more popular
- existence is not guaranteed ("being more popular" is not a linear order on matchings)

$$\begin{array}{llll} a_1 : & p_1 & p_2 & p_3 \\ a_2 : & p_1 & p_2 & p_3 \\ a_3 : & p_1 & p_2 & p_3 \end{array}$$

- characterize instances with popular matching
- decide existence and compute in time $O(\sqrt{n}m)$

SODA 05, joint work with D. Abraham, R. Irving, and T. Kavitha

Popular Matchings, Extensions

- Maestre (ICALP 2006): weigthed elements
- Manlove, Sng (ESA 2006): elements in B have capacities
- Mahdian (Conf. on Electronic Commerce): if $|B| > c \times |A|$ and preference lists are random, popular matchings exist
- Kavitha, Shah (ISAAC 2006): n^ω algorithm
- Abraham, Kavitha (SWAT 2006): for every matching N there is a popular matching M that is more popular than N.

Rank Maximal Matchings and Variants

- $r =$ maximal rank of any edge
- $s_i(M) =$ number of rank i edges in M
  - maximize the signature $(s_1, s_2, \ldots, s_r)$ — maximize happiness
  - maximize $(s_1 + s_2 + \ldots + s_r,\ s_1, s_2, \ldots, s_r)$ — max card, max happy
  - maximize $(s_1 + s_2 + \ldots + s_r,\ -s_r, -s_{r-1}, \ldots, -s_1)$ — max card, min unhappy

- first problem in time $O(\min(r \cdot n^{1/2} \cdot m, n \cdot m))$ and space $O(m)$

  SODA 04, joint work with R. Irving, T. Kavitha, D. Michail, and K. Paluch

- all problems in essentially this time and space

  unpublished, joint work with D. Michail

Strongly Stable Matchings

- both sides rank their incident edges (ties allowed)
- a matching M is stable if there is no blocking edge
- an edge $(x,y) \in E \setminus M$ is *blocking* if x would prefer to match up with y and y would not object, i.e.,
  - x prefers y over its current partner or is free
  - y prefers x over its current partner or is indifferent between them or is free
- decide existence of a stable matching and compute one
- we do so in time $O(nm)$, even if nodes in B have capacities
- previous best was $O(m^2)$ by R. Irving
- Irving's algorithm is used to match medical students and hospitals
- open problem: deal with couples

STACS 04, joint work with T. Kavitha, D. Michail, and K. Paluch

An Instance without a Stable Matching

$$\begin{array}{llcll} x_1: & w_1, w_2 & \qquad & w_1: & x_2, x_1 \\ x_2: & \{w_1, w_2\} & \qquad & w_2: & x_2, x_1 \end{array}$$

- both woman prefer x_2 to x_1.
- man x_1 prefers w_1 to w_2 and x_2 is indifferent between the women.
- every man ranks every woman and vice versa and hence any strongly stable matching must match all men and all women.
- no strongly stable matching exists
  - consider partner of x_1.
  - she prefers x_2 over x_1 and x_1 does not object

The Classical Algorithm (No Ties, Complete Instances)

MAX-PLANCK-GESELLSCHAFT

$M = \emptyset$;
while $\exists$ a free man m **do**
  let $e = (m, w)$ be the top choice of x;
  if w is free or prefers x over her current partner **then**
    dissolve the current marriage of w (if any) and add e to M;
  else
    discard e;
  end if
end while

- once matched, women stay matched   and to better and better partners
- alg constructs a complete and stable matching (man-optimal)
  - complete: every women is matched ultimately
  - if an engagement (m, w) is dissolved or rejected, it is not blocking with respect to the final matching
  - if an edge (m, w) is never considered, it is not ...
- alg for general case is similar, but more involved

Average Case Behavior of Matching Algs

Algorithms of Hopfcroft/Karp and Micali/Vazirani compute maximum cardinality matchings in bipartite or general graphs in time $O(\sqrt{n}m)$

observed behavior seems to be much better

number of phases seems to grow like $\log n$ ($n \le 10^6$ in experiments)

Motwani(JACM, 94): running time is $O(m \log n)$ with high probability for random graphs in the $G_{n,p}$ model provided that $p \geq (\ln n)/n$.

Our result: running time is $O(m \log n)$ with high probability for random graphs in the $G_{n,p}$ model provided that $p \geq c_0/n$.

$$c_0 = \begin{array}{rl} 9.6 & \text{for bipartite graphs} \\ 35.1 & \text{for general graphs} \end{array}$$

Open problem: what happens for p with $0 \le p \le c_0$?

Theory of Computing Systems 05, joint work with H. Bast, G. Schäfer, and H. Tamaki

Average Case Behavior II

- $G =$ random graph in $G_{n,p}$ model: every potential edge is present with probability p, independent of other edges.
- expected degree is pn for bipartite graphs, $p(n-1)$ for general graphs
- $p \geq c_0/n$,
- $c_0 = \begin{array}{rl} 9.6 & \text{for bipartite graphs} \\ 35.1 & \text{for general graphs} \end{array}$
- with high probability, G has the property that every non-maximum matching has an augmenting path of length $O(\log n)$
- algs of Hopcroft/Karp and Micali/Vazirani compute maximum matchings in expected time $O(m \log n)$

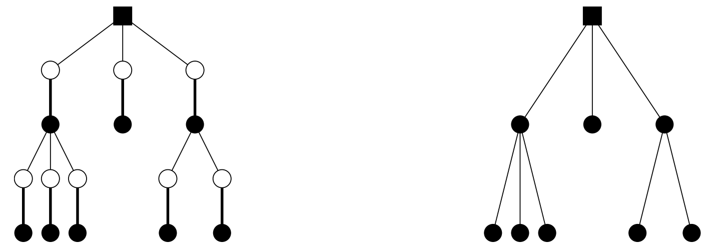
  because running time is $O(m \cdot L)$, where L is length of longest shortest augmenting path with respect to any non-maximum matching
- Motwani (JACM 94) proved the result for $p \geq (\ln n)/n$

Notation and Basic Facts

- $G = (V, E)$, graph
- matching = subset of edges no two of which share an endpoint
- maximum matching = matching of maximum cardinality
- $M \subseteq E$, matching
- matching edge = edge in M
- non-matching edge = edge outside M
- matched node = node incident to an edge in M
- free node = non-matched node
- alternating path $p = (e_1, e_2, \ldots, e_k)$ with $e_i \in M$ iff $e_{i+1} \notin M$
- augmenting path = alternating path connecting two free nodes
- if p is augmenting, $M \oplus p$ has one larger cardinality than M
- if M is non-maximum, there is augmenting path relative to it
- $S \subseteq V$, $\Gamma(S) =$ neighbors of the nodes in S

Motwani's Argument

- non-maximum matchings in expander graphs have short augmenting paths because alternating trees are bushy and hence reach all nodes after $\log n$ levels



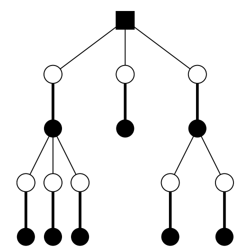
- expander graph: $|\Gamma(S)| \geq (1+\varepsilon)|S|$ for every node set S with $|S| \leq n/2$
- for $p \geq (\ln n)/n$: random graphs are essentially expander graphs

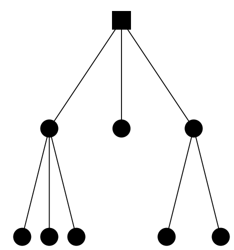
- sparse random graphs are far from being expander graphs
  - constant fraction of nodes is isolated
  - constant fraction of nodes has degree one
  - there are chains of length $O(\log n)$
  - nevertheless, our proof also uses the concept of expansion

Two Probabilistic Lemmas

An *alternating path tree* is a rooted tree of even depth, where each vertex in $Odd(T)$ has exactly one child.

We use $Even(T)$ to denote the nodes of even depth excluding the root. Then $|Odd(T)| = |Even(T)|$.





There are suitable constants ε, β, c_0 such that random graphs $G \in G(n,n,c/n)$, where $c \geq c_0$, with high probability have the following properties ($\varepsilon = 0.01$, $\beta = 2.6$, $c_0 = 9.6$ do):

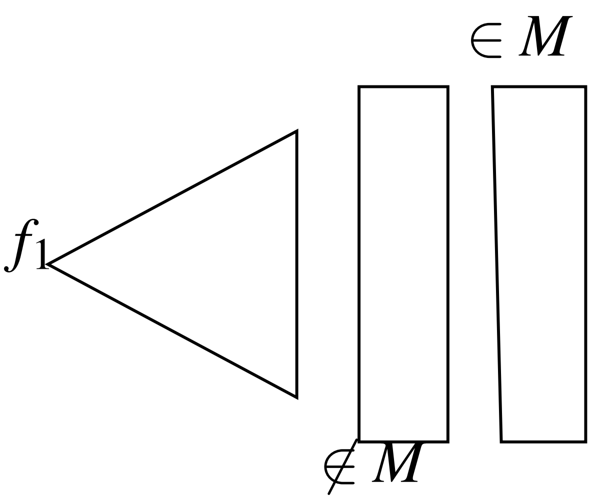
Lemma 1 (Expansion Lemma for Trees) *Each alternating path tree T with $\alpha \cdot \log n \leq |Even(T)| \leq n/\beta$ expands, i.e.,*

$$|\Gamma(Even(T))| \geq (1+\varepsilon) \cdot |Even(T)|$$

Lemma 2 (Large Sets Lemma) *Every two large disjoint sets of vertices, i.e., both of size at least n/β, have an edge between them.*

The Proof of the Main Theorem: Bipartite Case

- M non-maximum matching, p augmenting path, endpoints f_1 and f_2
- grow alternating trees T_1 and T_2 rooted at f_1 and f_2, respectively
  - suppose we have constructed even nodes at level $2j$
  - put their unreached neighbors into level $2j+1$
  - stop if one of the new nodes is free or belongs to other tree
  - put mates of new nodes into level $2j+2$
- grow the trees in phases: in each phase add two levels to both trees



Claim: process ends after a logarithmic number of phases

The Proof of the Main Lemma II

- if $|Even(T_i)| \geq n/\beta$ for $i = 1, 2$, the Large Sets Lemma guarantees an edge connecting them and the process stops
- Expansion Lemma implies that situation of preceding item is reached in a logarithmic number of phases
- Expansion Lemma guarantees expansion of trees with at least logarithmically many levels
- consider a phase $2j$ with $j \geq \alpha \log n$: then $|Even(T_i)| \geq \alpha \log n$
- assume $|Even(T_i)| < n/\beta$ and let T_i' be the next tree
- then $|Even(T_i')| \geq (1+\varepsilon) \cdot |Even(T_i)|$ and we have exponential growth

Why do Trees expand, if Sets do not?

Motwani used an expansion lemma for sets.

What is probability that some set does not expand, i.e., for some set S, $|S| = s$, we have $|T| \le \varepsilon s$ where $T = \Gamma(S) \setminus S$?

$$\sum_{t \le \varepsilon s} \binom{n}{s} \binom{n-s}{t} (1 - c/n)^{s(n-(s+t))}$$

- there are $\binom{n}{s}$ ways to choose S
- and $\binom{n-s}{t}$ ways to choose T
- and we want no edge from S to $V \setminus T$

Why do Trees expand, if Sets do not?

Motwani used an expansion lemma for sets.

What is probability that some set does not expand, i.e., for some set S, $|S| = s$, we have $|T| \le \varepsilon s$ where $T = \Gamma(S) \setminus S$?

$$\sum_{t \le \varepsilon s} \binom{n}{s} \binom{n-s}{t} (1 - c/n)^{s(n-(s+t))}$$

we concentrate on a single term and on the case where $s + t \ll n$. Then

$$\approx \left(\frac{en}{s}\right)^s \left(\frac{en}{t}\right)^t e^{-(c/n)sn} \approx \left(\frac{en}{s}\right)^s \left(\frac{en}{t}\right)^t e^{-cs}$$

we ignore the term involving t and obtain

$$\approx \left(\frac{en}{se^c}\right)^s$$

In order for this to be small one needs $c = \Omega(\log n)$.

And now for trees

- What happens if we require in addition that G contains a tree on S?
- We have an additional factor

$$s^{s-2}(c/n)^{s-1}$$

- the first factor counts the number of trees (Cayley's theorem)
- the second factor accounts for the fact that the edges of the tree must be present
- if we add this into our previous formula, we obtain

$$\approx \left(\frac{en}{se^c}\right)^s s^{s-2}(c/n)^{s-1} \le (n/c)s^2\left(\frac{ensc}{sne^c}\right)^s \le n^3\left(\frac{ec}{e^c}\right)^s$$

- and this is small if $s = \Omega(\log n)$ and c a sufficiently large constant:
  logarithmic size trees expand

- of course, the details are slightly more involved

Open Problems

- is the result true for all random graphs?
  - we need $c \geq c_0, \quad c_0 = 9.6$ for bipartite graphs, ...
  - result also holds for $c < 1$, since only logarithmic size connected components
  - what happens in between?

Summary

- Part I: One side ranks: notions of optimality
  - Rank-Optimal Matchings
  - Pareto-Optimal Matchings
  - Popular Matchings
- Part II: Both sides rank: stability
- Part III: Average Case Analysis of Matching Algorithms