How to Assign Papers to Referees Objectives, Algorithms, Open Problems

Kurt Mehlhorn
Max-Planck-Institut für Informatik
Saarbrücken
Germany

based on discussions with
N. Garg, T. Kavitha, A. Kumar, J. Mestre

Overview

- Motivation
- Informal Problem Definition
- Formal Problem Definition I
    - Formulation
    - An efficient algorithm
- Formal Problem Definition II
    - Formulation
    - Easy Cases
    - Hard Cases, NP-completeness and approximation

Slides are available at my home page

This is work in progress. We have more questions than answers

Motivation

I was program chair of ESA 2008.

After submission closes and before reviewing starts, the PC chair has to assign the papers to the PC members (called reviewers in the sequel).

What constitutes a good assignment?

Informal Problem Definition I

- n reviewers, i indexes reviewers
- m papers, j indexes papers
- v_{ij}, the valuation of paper j by reviewer i
  the interest of reviewer i in paper j
  the qualification of reviewer i for paper j
- the valuations can be determined in many different ways:
  - the PC chair invents them
  - papers and PC members provide key words, v_{ij} is a function of the number of common key words, e.g.,

$$v_{ij} = \frac{\text{number of common key words}}{\text{number of keywords provided by } i \text{ and } j}$$

  - reviewers provide values in $\{$NO, LOW, MEDIUM, HIGH$\}$
  - the last alternative is used by EasyChair (Andrei Voronkov), the system used for ESA 2008.

Informal Problem Definition II

MAX-PLANCK-GESELLSCHAFT

- n reviewers, i indexes reviewers m papers, j indexes papers
- bipartite graph $G = (\text{papers} \cup \text{reviewers}, E)$
- if $(i, j) \notin E$, j cannot review i conflict of interest
- for $(i, j) \in E$, v_{ij} is the value of assigning i to j.
- Objectives
  - each paper is reviewed at least k times
  - reviewers are not overloaded
  - papers are reviewed by qualified reviewers
  - reviewers get the papers that they are interested in
  - fairness
  - EasyChair converts the v_{ij} to numbers (LOW = 1, MEDIUM = 2, HIGH = 3) and computes a maximum weight assignment subject to the constraint that each paper is reviewed exactly k times (a number specified by the program chair) and the load among the reviewers is balanced.

Fairness, Quality of Review

reviewer 1: L L   reviewer 2: H H

is worse than

reviewer 1: L H   reviewer 2: L H

paper 1: L L   paper 2: H H

is worse than

paper 1: L H   paper 2: L H

Formalization I

- we view the assignment as proceeding in rounds:

revs	papers				
1	3	7	4	9	1
2	5	4	2	3	7
...					
n	3	1	4	7	9

revs	ranks (sorted)				
1	5	5	3	1	1
2	5	4	2	2	2
...					
n	3	1	1	1	1

- signature of a round:

  (# of rank r papers, # of rank $r-1$ papers, ..., # of rank 1 papers)

- objectives:
  - maximize signature of each round or
  - minimize reversed signature
  - both objectives maximize $\min_i$ # of rank r papers assigned to i

- can be reduced to a weighted b-matching problem

The Weights

- maximize signature of each round

- we aim for weights with the following properties:

- weights for a single round:

  - a paper of rank d contributes weight $(n+1)^d$ to the weight of a round

  - n rank $d-1$ assignments cannot make up for one rank d assignment

  - maximum weight of a round: $n(n+1)^r$, set $W = (n+1)^{r+1}$

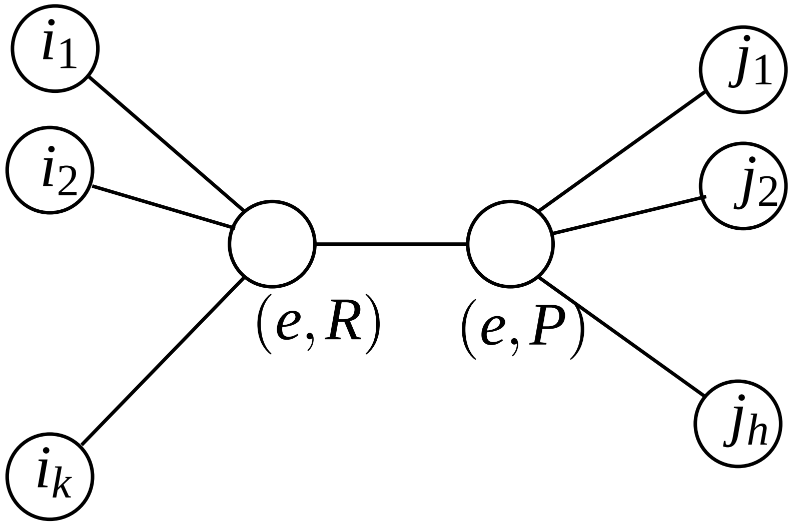
- total weight of assignment = $w_1 W^k + w_2 W^{k-1} + \ldots + w_0 W^0$

  $w_\ell =$ weight of round ℓ and k is the number of rounds

- $(x+1) \cdot W^d > x \cdot W^d + (W-1) \cdot W^{d-1} + \ldots + (W-1) \cdot W^0$

The Weighted Bipartite Matching Problem

- vertex i_ℓ represents reviewer i in round ℓ, $1 \le \ell \le k$
- vertex j_c represents copy c of paper j, $1 \le c \le h := nk/m$
- for each edge $e = (i, j)$ of rank d, we have two vertices (e, R) and (e, P) and the following edges



- i_ℓ is connected to (e, R) and has weight $(n+1)^d W^{n+1-\ell}$, $1 \le \ell \le k$
- (e, R) is connected to (e, L) and has weight 0
- (e, P) is connected to j_c and has weight 0, $1 \le c \le h$.
- If j is assigned to i in round ℓ, $(i_\ell, (e, R))$ and $((e, P), j_c)$ are in M.
- If j is not assigned to i in any round, then $((e, R), (e, P)) \in M$.
- weight of assignment = weight of matching

Remarks

- same approach works with roles of papers and reviewers reversed
- signature of a reviewer = sorted sequence of ranks of assigned papers
- Open Problem: maximize the minimal signature
- for two ranks, say Low and High: maximize the number of H's in each round

H	H	H	H	H	H	H	H	L	L	L
H	H	H	H	H	H	L	L	L	L	L
H	H	H	H	H	H	L	L	L	L	L
H	H	H	H	H	L	L	L	L	L	L
H	H	L	L	L	L	L	L	L	L	L

- this maximizes the minimum signature and subject to this maximizes the second smallest signature and subject to this . . .

Formalization II

- inspired by allocation of indivisible goods (Santa Claus problem)
- sources
  - Bezakova, Dani: ACM SIGecom 2005
  - Lenstra, Schmoys, Tardos: Math Program. 1990
- the values v_{ij} are numbers and it makes sense to add them
- binary variables x_{ij} with $x_{ij} = 1$ iff paper j is assigned to reviewer i
$L_i = \sum_j x_{ij}$	load of reviewer i
$I_i = \sum_j v_{ij} x_{ij}$	total value of assignment for reviewer i
$L_j = \sum_i x_{ij}$	number of reviews for paper j
$I_j = \sum_i v_{ij} x_{ij}$	total value of assignment for paper j

Formalization II

- the valuations v_{ij} are numbers and it makes sense to add them
- binary variable x_{ij} with $x_{ij} = 1$ iff paper j is assigned to reviewer i
- $L_i = \sum_j x_{ij}$ load of reviewer i
  $I_i = \sum_j v_{ij} x_{ij}$ total value of assignment for reviewer i
  $L_j = \sum_i x_{ij}$ number of reviews for paper j
  $I_j = \sum_i v_{ij} x_{ij}$ total value of assignment for paper j
- Constraints
  $L_i \leq h$ or $L_i = h$ bound on reviewer load
  $I_i \geq s$ lower bound on reviewer value
  $L_j = k$ or $L_j \geq k$ bound on number of reviews per paper
  $I_j \geq t$ lower bound on total interest level for any paper
- we obtain different variants according to which constraints have to be enforced and according to which quantity is to be optimized

Easy Cases

MAX-PLANCK-GESELLSCHAFT

- Constraints

$L_i \leq h$ or $L_i = h$	bound on reviewer load
$I_i \geq s$	lower bound on reviewer value
$L_j = k$ or $L_j \geq k$	bound on number of reviews per paper
$I_j \geq t$	lower bound on total interest level for any paper

- only load constraints: a standard b-matching problem

- only constraints on either papers or reviewers, but not on both, e.g.,

  maximize t subject to $L_j = k$ for all papers j (k is set by PC chair) and $I_j \geq t$ for all papers j

  - for each paper select the k reviewers that show the largest interests for this paper.
  - remark: load for reviewers may be highly unbalanced

Hard Cases?

- constraints on reviewers and papers and not just load constraints
- some versions are known to be NP-complete
- if only one kind of constraint for reviewers and only one kind of constraint for papers, e.g.,

  maximize t subject to $L_i = h$ for all reviewers i (h is set by PC chair) and $I_j \geq t$ for all papers j

  we know a good approximation algorithm
- constraints for both and more than kind of constraint for either papers or reviewers, e.g.,

  maximize t subject to $L_i = h$ for all reviewers i, $L_j = nh/m$ for all papers j, and $I_j \geq t$ for all papers j

  we know nothing beyond maybe NP-completeness

One Load, One Interest Constraint: Hardness

MAX-PLANCK-GESELLSCHAFT

- maximize t subject to $L_j = k$ for all papers j (k is set by PC chair) and $I_i \geq t$ for all reviewers i
- problem is NP-complete
  - 2 reviewers, $2n$ papers, $k = 1$, $v_{1j} = v_{2j}$ for all j
  - solution with $t = \sum_j v_{1j}/2$ exists iff subset problem ...
- Open Problems
  - valuations are small integers, say in $\{1, 2, 3\}$
  - maximize t subject to $L_i = h$ for all reviewers i and $I_j \geq t$ for all papers j

One Load, One Interest Constraint: Approximation

- maximize t subject to $L_j = k$ for all papers j (k is set by PC chair) and
  $I_i \geq t$ for all reviewers i

- let (x^*_{ij}) and t_{opt} be an optimal solution to the linear program

  maximize t subj. to $\sum_i x_{ij} = k$ for all j, $\sum_j v_{ij} x_{ij} \geq t$ for all i, $0 \leq x_{ij} \leq 1$

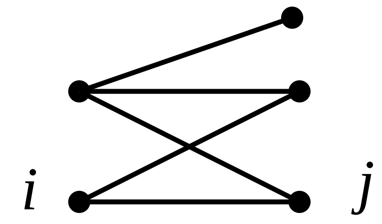
- Claim: (x^*_{ij}) can be rounded to an integer solution with $t \geq t_{opt} - v_{max}$, where $v_{max} = \max_{ij} v_{ij}$

- same claim holds for

  maximize t subject to $L_i = h$ for all reviewers i (h is set by PC chair)
  and $I_j \geq t$ for all papers j

The Graph of Fractional Variables I

- consider the bipartite graph defined by the fractional variables, i.e., the x^*_{ij} with $0 < x^*_{ij} < 1$



- Claim: a connected component with ℓ vertices contains at most ℓ edges

- the fractional variables are defined by a system of equations from among $\sum_i x_{ij} = k$ for $j \in \{1, n\}$ and $\sum_j v_{ij} x_{ij} = t_{opt}$ for some $i \in \{1, n\}$

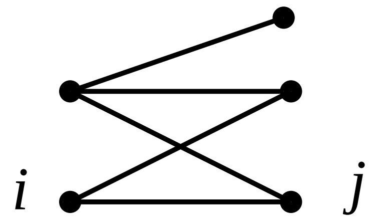
- consider a connected component involving ℓ vertices

- only the ℓ equations corresponding to these vertices talk about the (variables corresponding to the) edges between these vertices

- thus at most ℓ non-zero variables in basic feasible solution

The Graph of Fractional Variables II

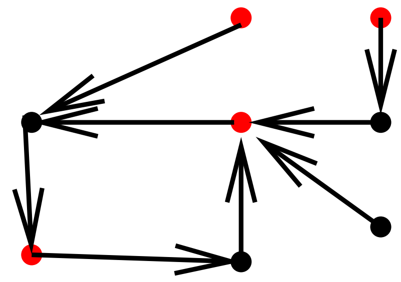
- consider the bipartite graph defined by the x^*_{ij} with $0 < x^*_{ij} < 1$



Claim: a connected component with ℓ vertices contains at most ℓ edges

- each connected component is a tree plus one edge

  orient the cycle arbitrarily and orient the edges in the trees towards the cycle



- for each paper, the incident fractional variables sum to an integer
- round the outgoing edge plus the right number of incoming edges to one; this gurantees $L_j = k$ for all j
- for each reviewer i: all incoming edges are rounded to one and at most the outgoing edge is rounded to zero; this guarantees $I_i \geq t_{opt} - v_{max}$

Two Interest Constraints: Approximation

MAX-PLANCK-GESELLSCHAFT

- minimize $\sum_{ij} x_{ij}$ subj. to $I_i \geq s$ for every reviewer i and $I_j \geq t$ for all papers j
- let (x^*_{ij}) be a basic feasible solution to the linear program, let $R^* = \sum_{ij} x^*_{ij}$

  minimize $\sum_{ij} x_{ij}$ subj. to $\sum_j v_{ij} x_{ij} \geq s$ for all i and $\sum_j v_{ij} x_{ij} \geq t$ for all i, $0 \leq x_{ij} \leq 1$
- Claim: (x^*_{ij}) can be rounded to an integer solution with $I_i \geq s - v_{max}$ and $I_j \geq t - v_{max}$ for all i and j and total number of reviews less $R^* + n$.
- Rounding Rule: for a reviewer i, let r_i be the sum of the fractional variables incident to i.

  round the outgoing edge to 1, and the $\lceil r_i - 1 \rceil$ incoming edges of largest valuation

One Load, One Interest Constraint: Refinement

MAX-PLANCK-GESELLSCHAFT

- every (fractional) solution to $L_j = k$ for all papers j gives rise to a vector $sort(I_1, I_2, \ldots, I_n)$

  let $(t_1^*, \ldots, t_n^*)$ be the optimal (lexicographically largest) vector over all fractional solutions

  on can obtain an integer solution with

$$sort(I_1, \ldots, I_n) \geq (t_1^* - v_{max}, \ldots, t_n^* - v_{max})$$

- solution
  - determine optimal fractional solution and then round as above
  - the following LP determines t_ℓ^*

    maximize t_ℓ^* subject to

    $L_j = k$ for all j and

    $\sum_{i \in R_q} I_i \geq t_1^* + \ldots t_q^*$ for all $q \leq \ell$ and all set R_q of reviewers of size q

What Next?

- What are the right objectives?
- Which objectives are easy, which are hard?
- Approximation algs for the hard objectives
- Exact algorithms for the hard objectives
- Experiments
- Incorporation into EasyChair