Aggregating partial rankings with applications to peer grading in MOOCs

Alexandros A. Voudouris

joint work with

Ioannis Caragiannis and George A. Krimpas

University of Patras and CTI "Diophantus"

About MOOCs

- **Massive**: available to a large number of people
  - For example, Coursera has ~12 million users (March 2015)
- **Online**: through the Internet
  - Easy access from everywhere
- **Open**: (usually) no cost for the students
- **Courses**: series of lectures on a subject
  - High level education from top universities

The challenge

Evaluate the performance of this huge number of students in an examination.

- **Easy solution:** use *closed* type questions (like multiple choice) that can be evaluated automatically by a computer
- But… there are courses where students must be examined in *open* type questions!
  - E.g., can they solve a mathematical exercise or write an essay?
- Such questions need to be evaluated by human graders
  - Unfortunately, AI is not that advanced yet …
- The available qualified human resources are limited and extremely costly

Overcoming the problem

Peer assessments

In many courses, the most meaningful assignments cannot be easily graded by a computer. That's why we use peer assessments, where learners can evaluate and provide feedback on each other's work. This technique has been shown in many studies to result in accurate feedback for the learner and a valuable learning experience for the grader.

coursera

Peer grading

Each student must grade a small number of other students' exam papers. [Piech et al. (2013)]

Ordinal peer grading

- Students grading with absolute numerical scores can be tricky…
  - They are not professional graders ⇒ they will do mistakes
  - They may act strategically ⇒ assign low scores to optimize their relative performance (is this possible? can we avoid it?)

Ordinal peer grading

- Each student *orders* the exam papers he is given from the best to the worst ⇒ *partial rankings*
- The partial rankings are *aggregated* into a global ranking that represents the students' relative performance [Shah et al. (2013), Raman and Joachims (2014)]

Important questions

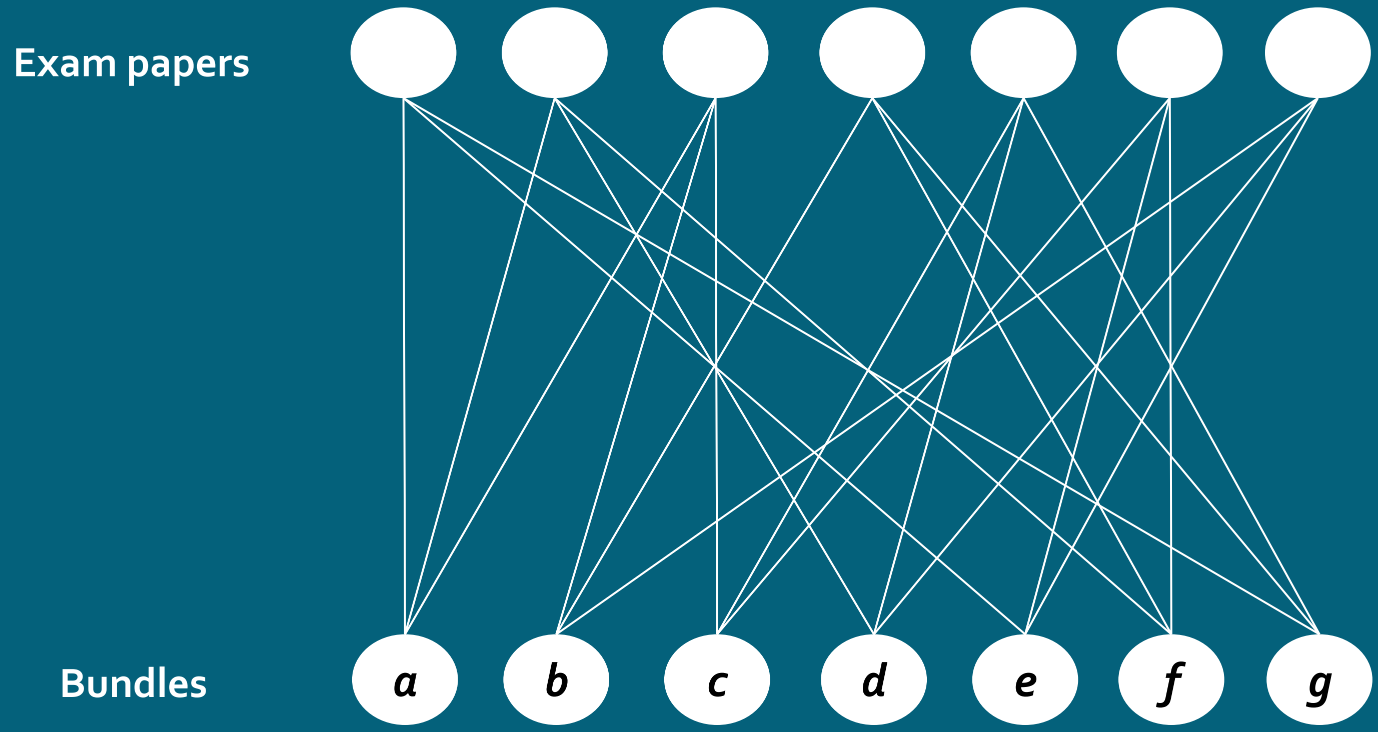
- **How many exam papers should we give to each student?**
  - k exam papers per student

- **How should we distribute the exam papers?**
  - Every exam paper is given to k students
- **How should we aggregate the partial rankings into a global one?**
  - The global ranking should be as close as possible to the objective global one (the *ground truth*)
  - We can think of the ground truth as the global ranking a professional grader would come up with, provided full information

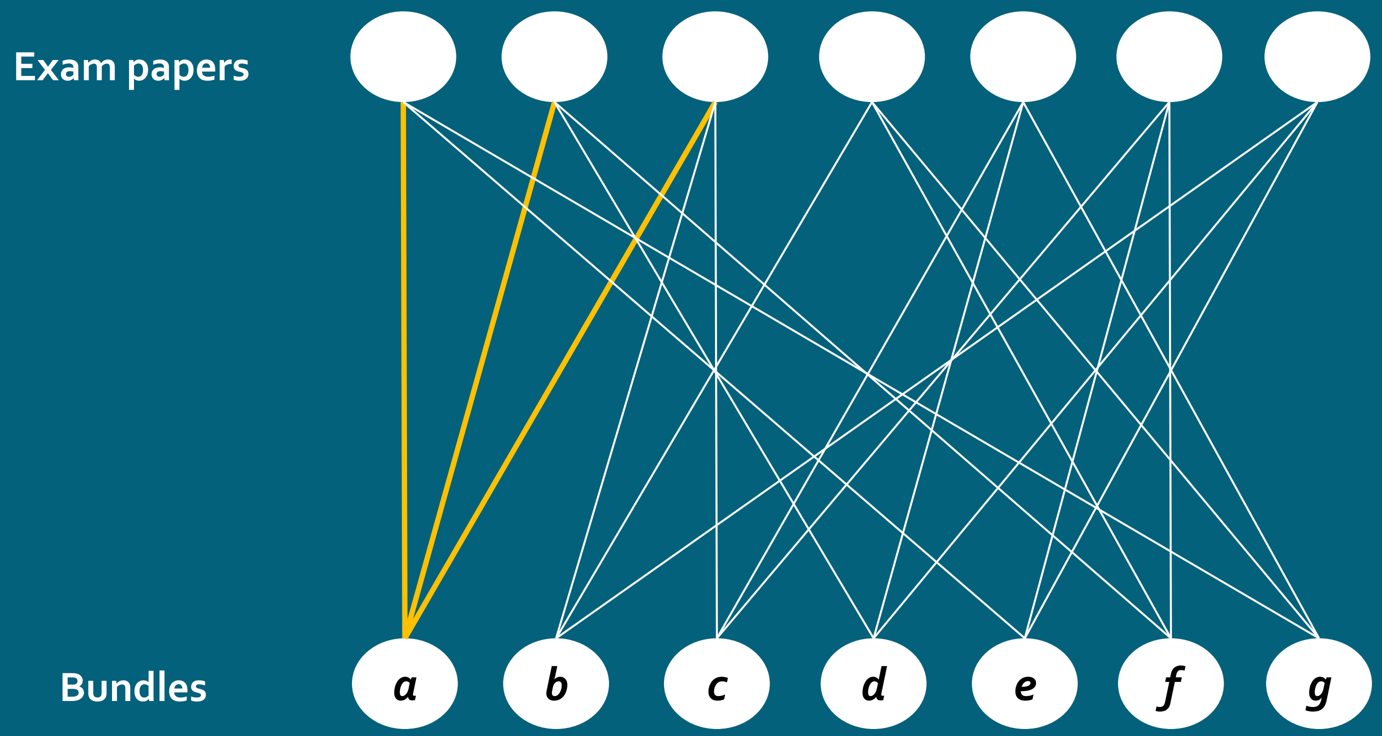
The model

- n students
- k exam papers per student
- (n, k)-grading scheme: a collection of *n bundles* each containing k exam papers
  - It defines the way exam papers are distributed to graders
- A grading scheme is represented by a bipartite k-regular graph $G = (U, V, E)$ called *bundle graph*
  - U = set of exam papers, V = set of bundles
  - An edge $e = (u, v)$ represents the fact that exam paper u belongs in bundle v
- Bundles are allocated to students such that no student is allocated a bundle containing his own exam paper

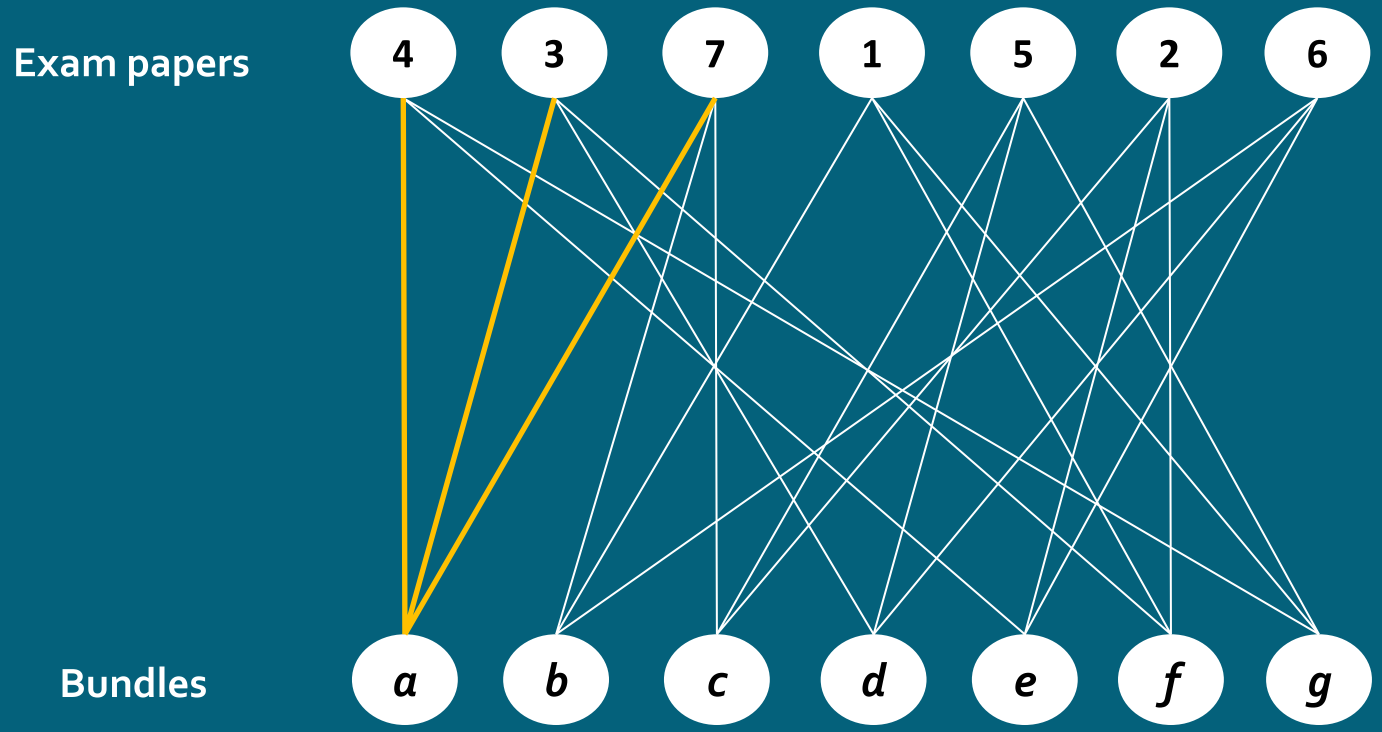
Example: (7,3)-grading scheme



Example: (7,3)-grading scheme



Example: (7,3)-grading scheme



Example: (7,3)-grading scheme + Borda

bundles	
a	3 4 7
b	1 6 7
c	2 5 7
d	3 5 6
e	2 4 6
f	1 2 3
g	1 4 5

Example: (7,3)-grading scheme + Borda

students
1
2
3
4
5
6
7

bundles	
a	3 4 7
b	1 6 7
c	2 5 7
d	3 5 6
e	2 4 6
f	1 2 3
g	1 4 5

Example: (7,3)-grading scheme + Borda

students	bundles	
1	***a***	3 4 7
2	***b***	1 6 7
3	***c***	2 5 7
4	***d***	3 5 6
5	***e***	2 4 6
6	***f***	1 2 3
7	***g***	1 4 5

Example: (7,3)-grading scheme + Borda

students
1
2
3
4
5
6
7

bundles	
a	3 4 7
b	1 6 7
c	2 5 7
d	3 5 6
e	2 4 6
f	1 2 3
g	1 4 5

partial rankings
$3 \succ 4 \succ 7$
$1 \succ 6 \succ 7$
$2 \succ 5 \succ 7$
$3 \succ 6 \succ 5$
$2 \succ 6 \succ 4$
$3 \succ 1 \succ 2$
$1 \succ 4 \succ 5$

Example: (7,3)-grading scheme + Borda

students	bundles		partial rankings	Borda
1	***a***	3 4 7	$3 \succ 4 \succ 7$	8
2	***b***	1 6 7	$1 \succ 6 \succ 7$	7
3	***c***	2 5 7	$2 \succ 5 \succ 7$	9
4	***d***	3 5 6	$3 \succ 6 \succ 5$	5
5	***e***	2 4 6	$2 \succ 6 \succ 4$	4
6	***f***	1 2 3	$3 \succ 1 \succ 2$	6
7	***g***	1 4 5	$1 \succ 4 \succ 5$	3

Example: (7,3)-grading scheme + Borda

students	bundles		partial rankings	Borda
1	***a***	3 4 7	$3 \succ 4 \succ 7$	8
2	***b***	1 6 7	$1 \succ 6 \succ 7$	7
3	***c***	2 5 7	$2 \succ 5 \succ 7$	9
4	***d***	3 5 6	$3 \succ 6 \succ 5$	5
5	***e***	2 4 6	$2 \succ 6 \succ 4$	4
6	***f***	1 2 3	$3 \succ 1 \succ 2$	6
7	***g***	1 4 5	$1 \succ 4 \succ 5$	3

Example: (7,3)-grading scheme + Borda

students	bundles		partial rankings	Borda
1	***a***	3 4 7	$3 \succ 4 \succ 7$	8
2	***b***	1 6 7	$1 \succ 6 \succ 7$	7
3	***c***	2 5 7	$2 \succ 5 \succ 7$	9
4	***d***	3 5 6	$3 \succ 6 \succ 5$	5
5	***e***	2 4 6	$2 \succ 6 \succ 4$	4
6	***f***	1 2 3	$3 \succ 1 \succ 2$	6
7	***g***	1 4 5	$1 \succ 4 \succ 5$	3

Example: (7,3)-grading scheme + Borda

students	bundles		partial rankings	Borda
1	***a***	3 4 7	$3 \succ 4 \succ 7$	8
2	***b***	1 6 7	$1 \succ 6 \succ 7$	7
3	***c***	2 5 7	$2 \succ 5 \succ 7$	9
4	***d***	3 5 6	$3 \succ 6 \succ 5$	5
5	***e***	2 4 6	$2 \succ 6 \succ 4$	4
6	***f***	1 2 3	$3 \succ 1 \succ 2$	6
7	***g***	1 4 5	$1 \succ 4 \succ 5$	3

global ranking
$3 \succ 1 \succ 2 \succ 6 \succ 4 \succ 5 \succ 7$

Example: (7,3)-grading scheme + Borda

students	bundles		partial rankings	Borda
1	***a***	3 4 7	$3 \succ 4 \succ 7$	8
2	***b***	1 6 7	$1 \succ 6 \succ 7$	7
3	***c***	2 5 7	$2 \succ 5 \succ 7$	9
4	***d***	3 5 6	$3 \succ 6 \succ 5$	5
5	***e***	2 4 6	$2 \succ 6 \succ 4$	4
6	***f***	1 2 3	$3 \succ 1 \succ 2$	6
7	***g***	1 4 5	$1 \succ 4 \succ 5$	3

global ranking	ground truth
$3 \succ 1 \succ 2 \succ 6 \succ 4 \succ 5 \succ 7$	$1 \succ 2 \succ 3 \succ 4 \succ 5 \succ 6 \succ 7$

Example: (7,3)-grading scheme + Borda

students	bundles		partial rankings	Borda
1	***a***	3 4 7	$3 \succ 4 \succ 7$	8
2	***b***	1 6 7	$1 \succ 6 \succ 7$	7
3	***c***	2 5 7	$2 \succ 5 \succ 7$	9
4	***d***	3 5 6	$3 \succ 6 \succ 5$	5
5	***e***	2 4 6	$2 \succ 6 \succ 4$	4
6	***f***	1 2 3	$3 \succ 1 \succ 2$	6
7	***g***	1 4 5	$1 \succ 4 \succ 5$	3

global ranking	ground truth
$3 \succ 1 \succ 2 \succ 6 \succ 4 \succ 5 \succ 7$	$1 \succ 2 \succ 3 \succ 4 \succ 5 \succ 6 \succ 7$

Example: (7,3)-grading scheme + Borda

students	bundles		partial rankings	Borda
1	***a***	3 4 7	$3 \succ 4 \succ 7$	8
2	***b***	1 6 7	$1 \succ 6 \succ 7$	7
3	***c***	2 5 7	$2 \succ 5 \succ 7$	9
4	***d***	3 5 6	$3 \succ 6 \succ 5$	5
5	***e***	2 4 6	$2 \succ 6 \succ 4$	4
6	***f***	1 2 3	$3 \succ 1 \succ 2$	6
7	***g***	1 4 5	$1 \succ 4 \succ 5$	3

global ranking	ground truth
$3 \succ 1 \succ 2 \succ 6 \succ 4 \succ 5 \succ 7$	$1 \succ 2 \succ 3 \succ 4 \succ 5 \succ 6 \succ 7$

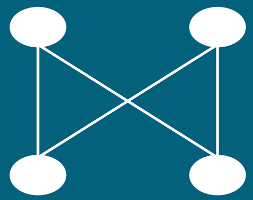
- 16/21 (=76.1%) correctly pairwise relations

Grading scenarios

- **Perfect grading:** the partial rankings are *consistent* to the ground truth
  - After all students have submitted their exam papers, the instructor announces indicatives solutions and grading instructions
  - The students use this info when grading
- **Imperfect grading:** the partial rankings are not necessarily perfect
  - No info is given by the instructor
  - Students' grading performance depends on the level they have studied for the exam

Bundle graph types

- Type 1: **Random graphs**
  - Constructed by considering $K_{n,n}$ and randomly picking k disjoint perfect matchings
- Type 2: **Square-free graphs**
  - Consist of multiple disjoint copies of a small **square-free** bipartite graph



- Type 3: **Disjoint Copies of $K_{k,k}$**

- We use highly disconnected bundle graphs in order to challenge our methods

Borda as an aggregation method

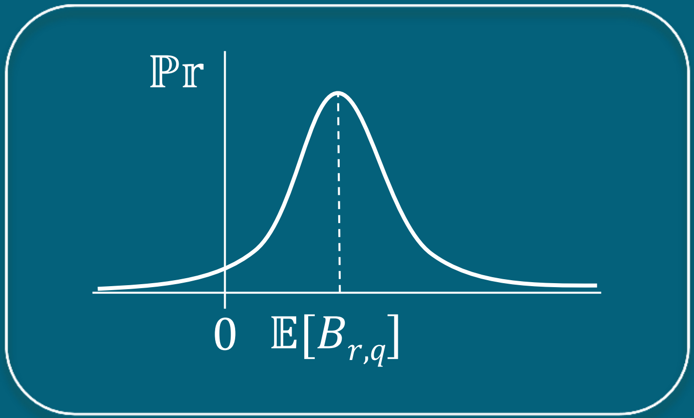
- Students order the exam papers in their bundles
  $\Rightarrow$ student i produces a partial ranking $(a_1, a_2, \ldots, a_k)$
- Exam paper a_j is assigned k-j+1 points
- The Borda score of a student equals the total points that his exam paper is assigned

Theorem

Under perfect grading, the expected fraction of pairwise relations correctly recovered by Borda is at least 1-$\mathcal{O}(1/k)$ in the case where the bundle graph is square-free, and at least 1-$\mathcal{O}\left(1/\sqrt{k}\right)$ in general

Proof idea

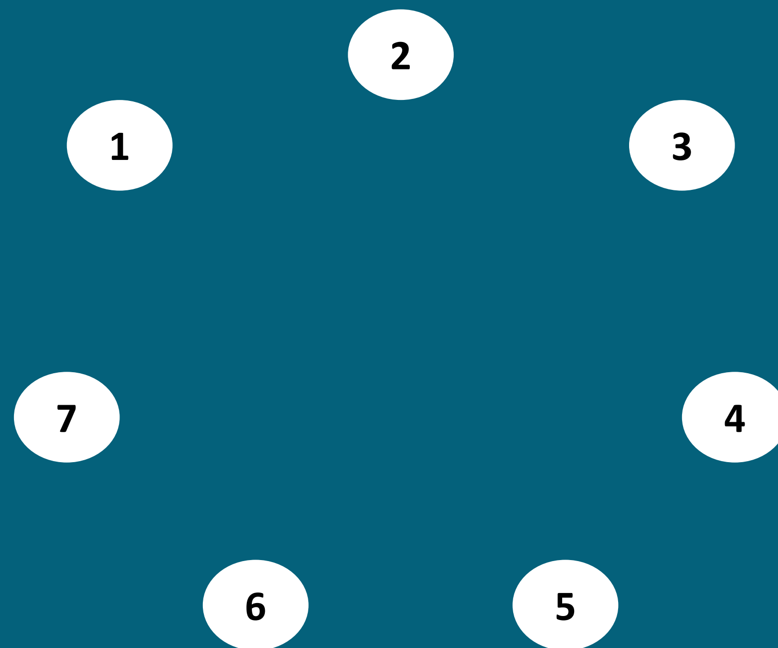
- Two exam papers x and y with ranks $r<q$ in the ground truth
- $\mathbb{E}[B_{r,q}]$ is proportional to $q-r$
- $B_{r,q}$ is sharply concentrated around $\mathbb{E}[B_{r,q}]$
  - Martingales and Azuma-Hoeffding inequality



- $\mathbb{Pr}[x \succ y] = \mathbb{Pr}[B_{r,q} > 0]$
- The bound follows by summing over all the (r, q) pairs

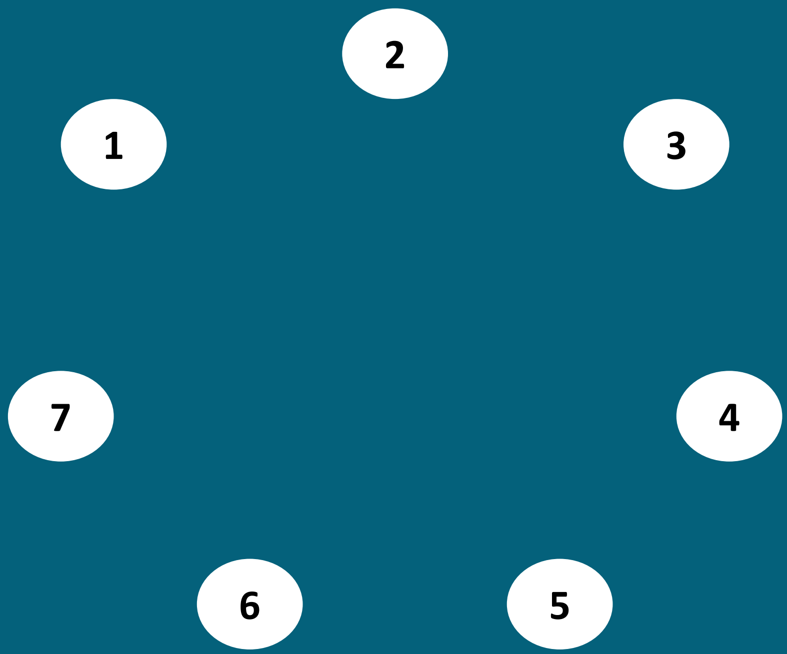
Random Serial Dictatorship (RSD)

partial rankings
$3 \succ 5 \succ 6$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$
$7 \succ 1 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 4 \succ 5$
$4 \succ 3 \succ 7$



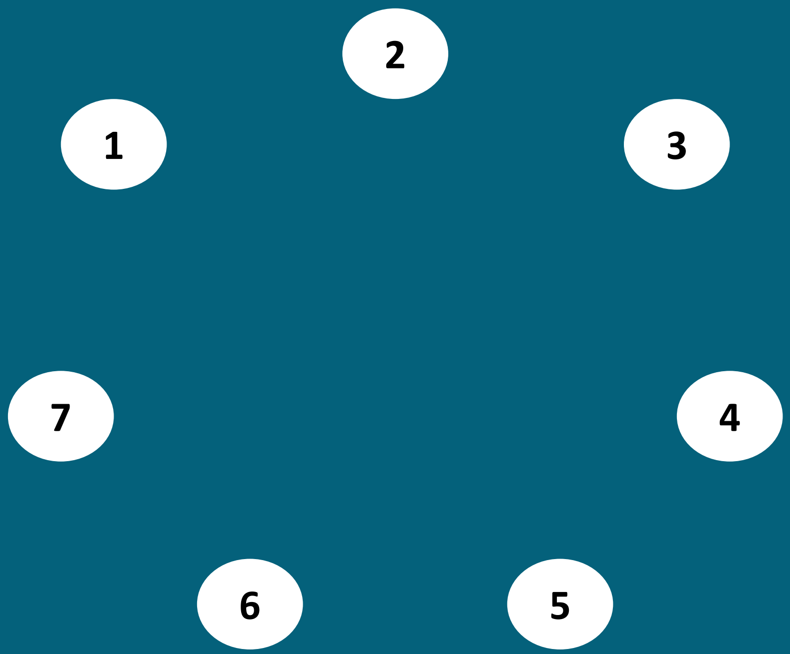
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



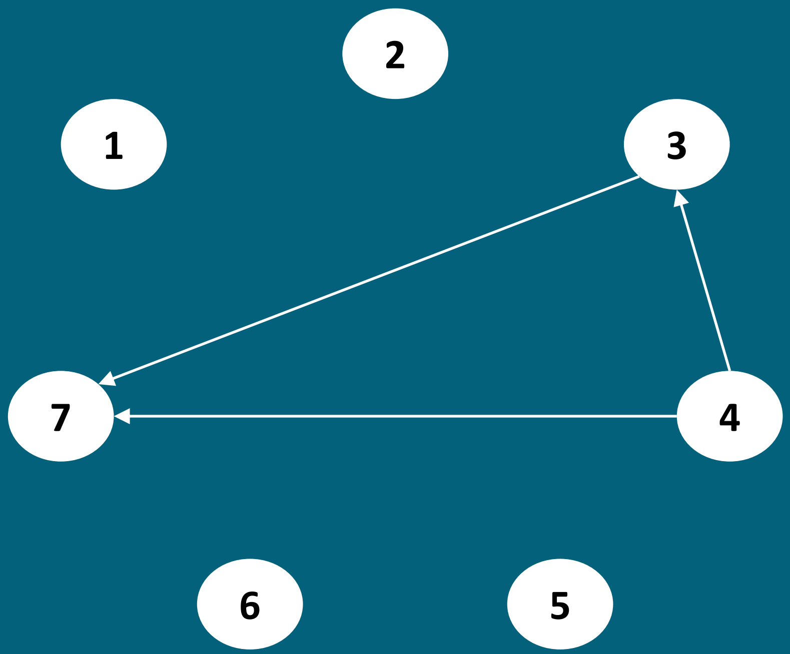
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



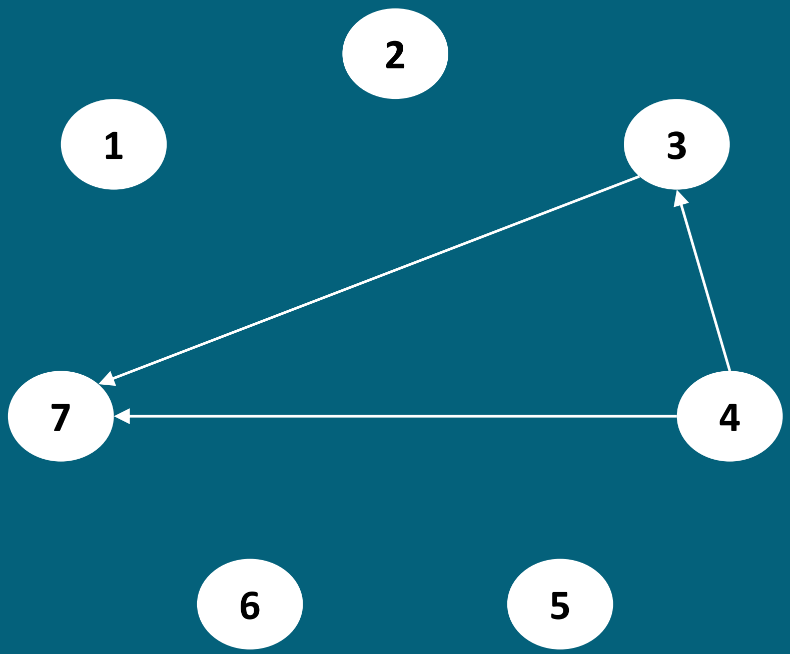
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



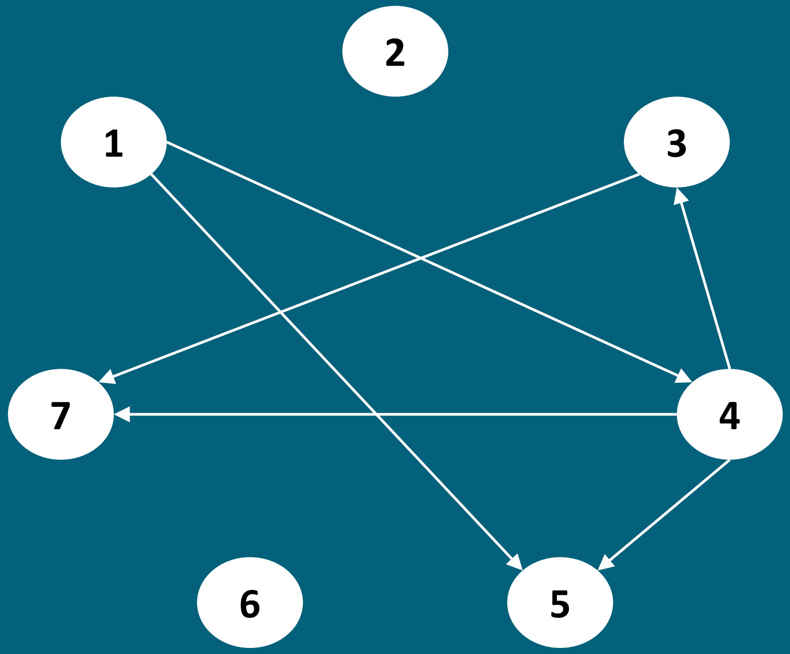
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



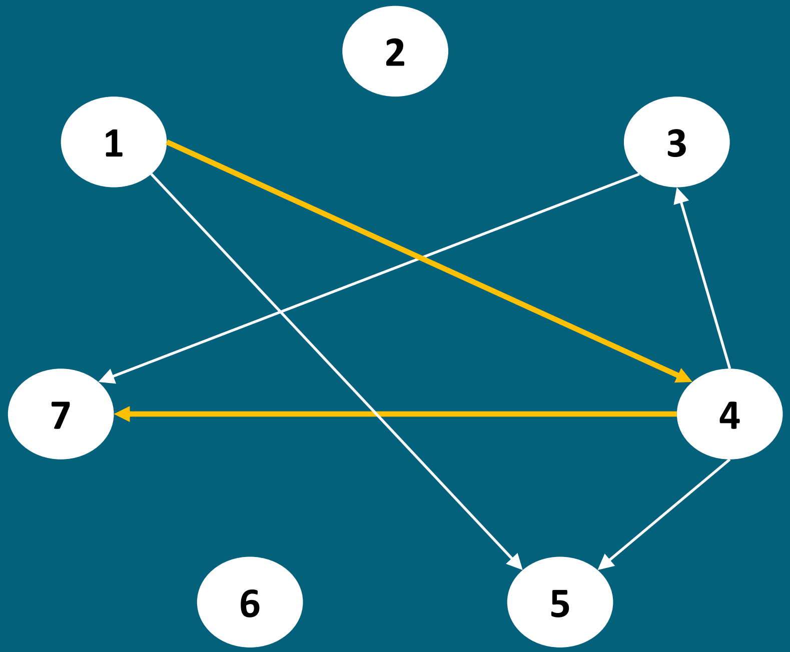
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



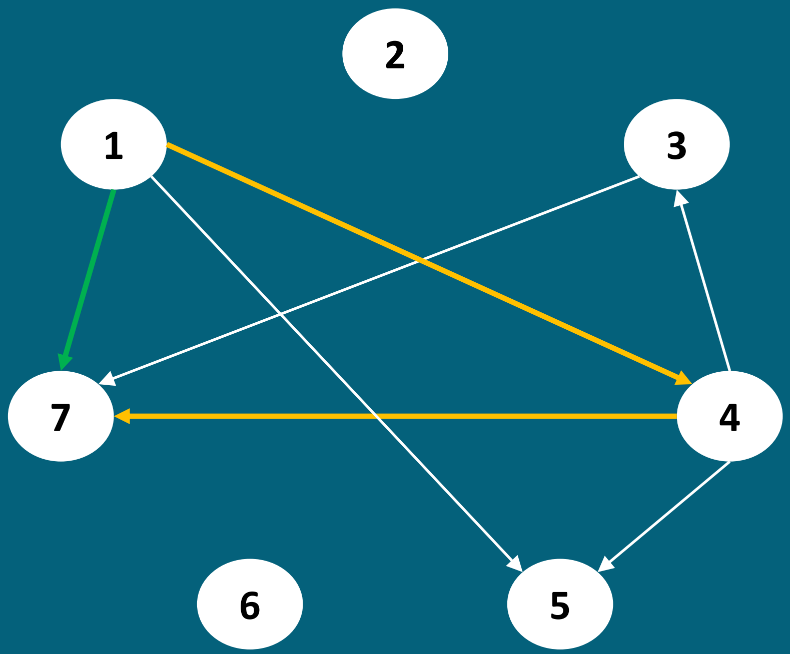
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



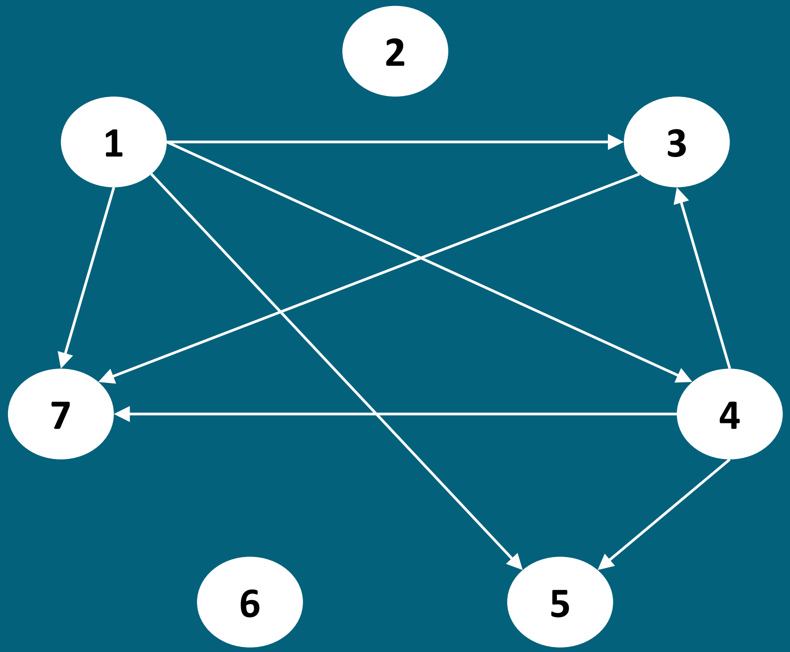
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



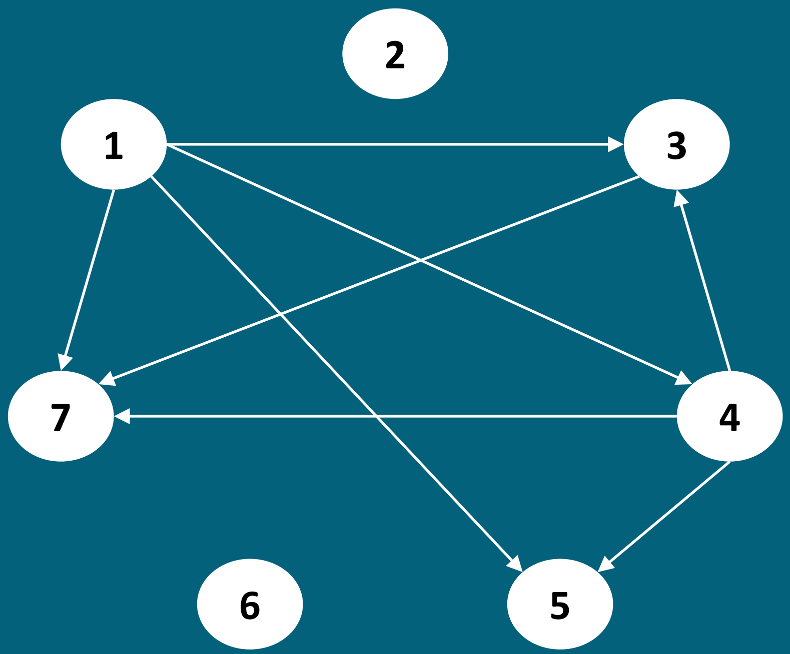
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



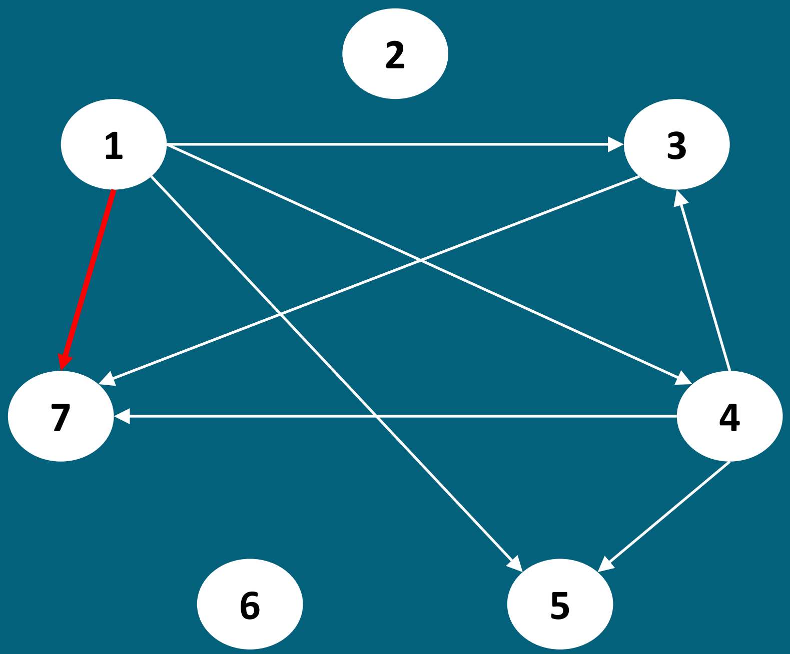
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



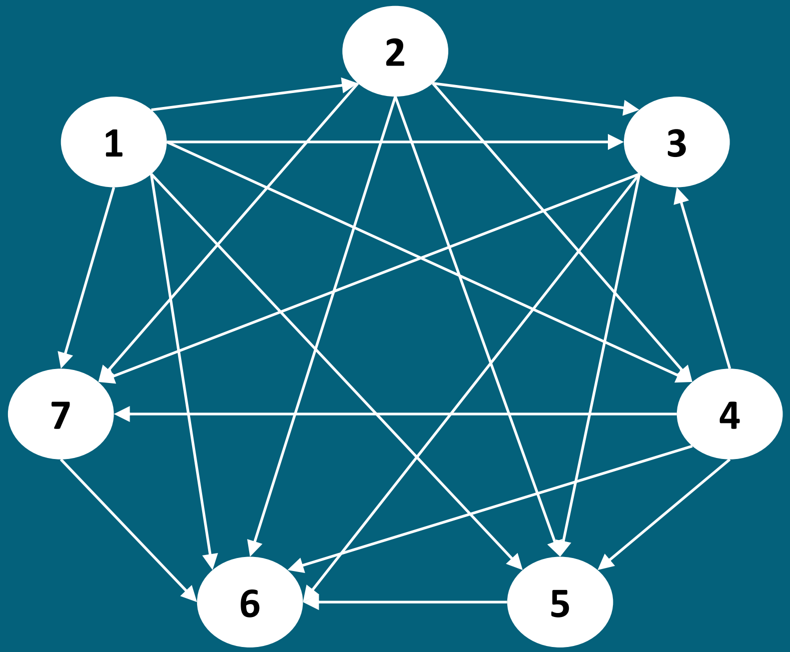
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



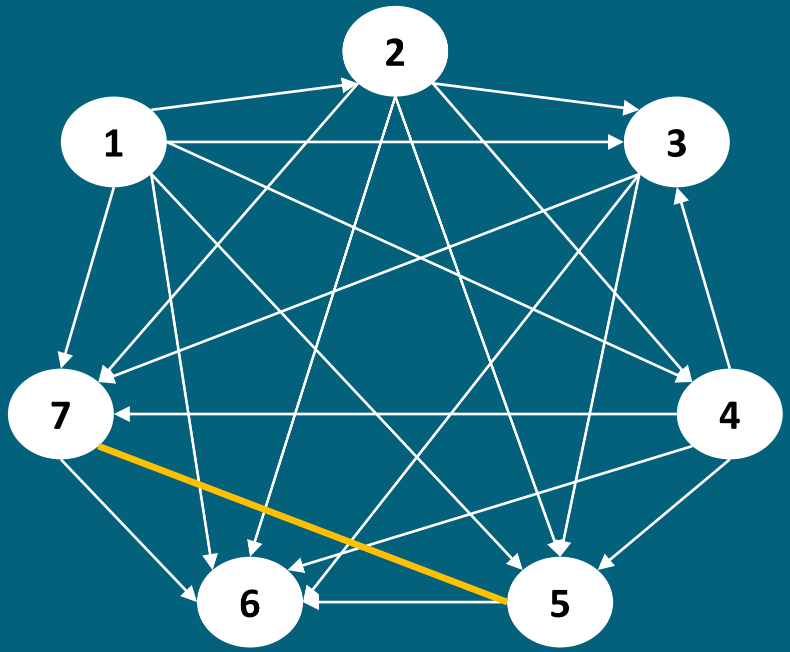
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



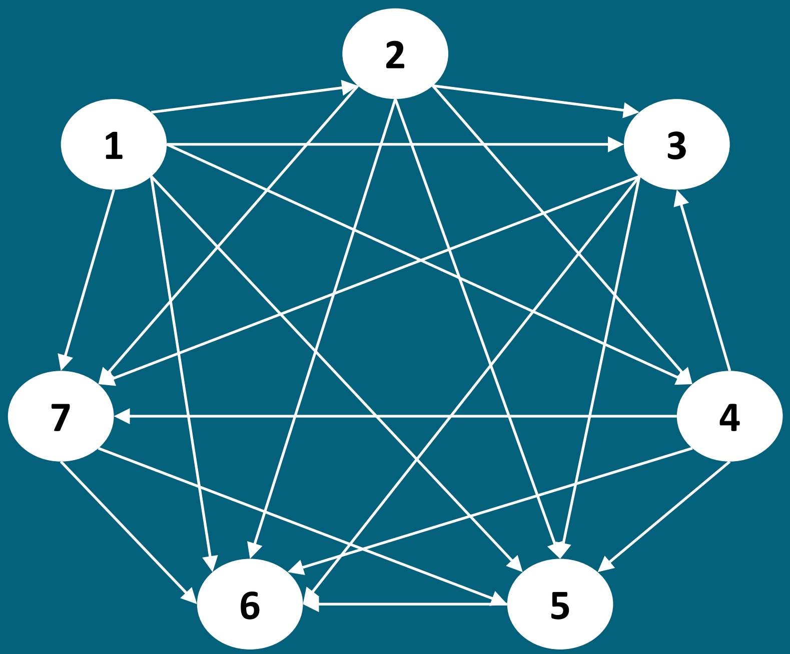
Random Serial Dictatorship (RSD)

partial rankings
4 ≻ 3 ≻ 7
1 ≻ 4 ≻ 5
7 ≻ 1 ≻ 6
3 ≻ 5 ≻ 6
2 ≻ 6 ≻ 5
1 ≻ 2 ≻ 7
2 ≻ 3 ≻ 4



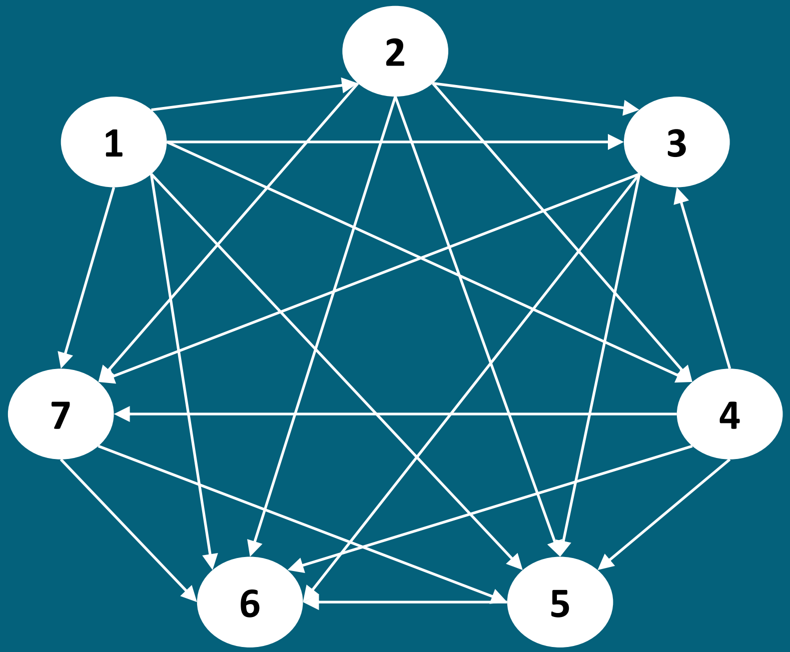
Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



Random Serial Dictatorship (RSD)

partial rankings
$4 \succ 3 \succ 7$
$1 \succ 4 \succ 5$
$7 \succ 1 \succ 6$
$3 \succ 5 \succ 6$
$2 \succ 6 \succ 5$
$1 \succ 2 \succ 7$
$2 \succ 3 \succ 4$



global ranking
$1 \succ 2 \succ 4 \succ 3 \succ 7 \succ 5 \succ 6$

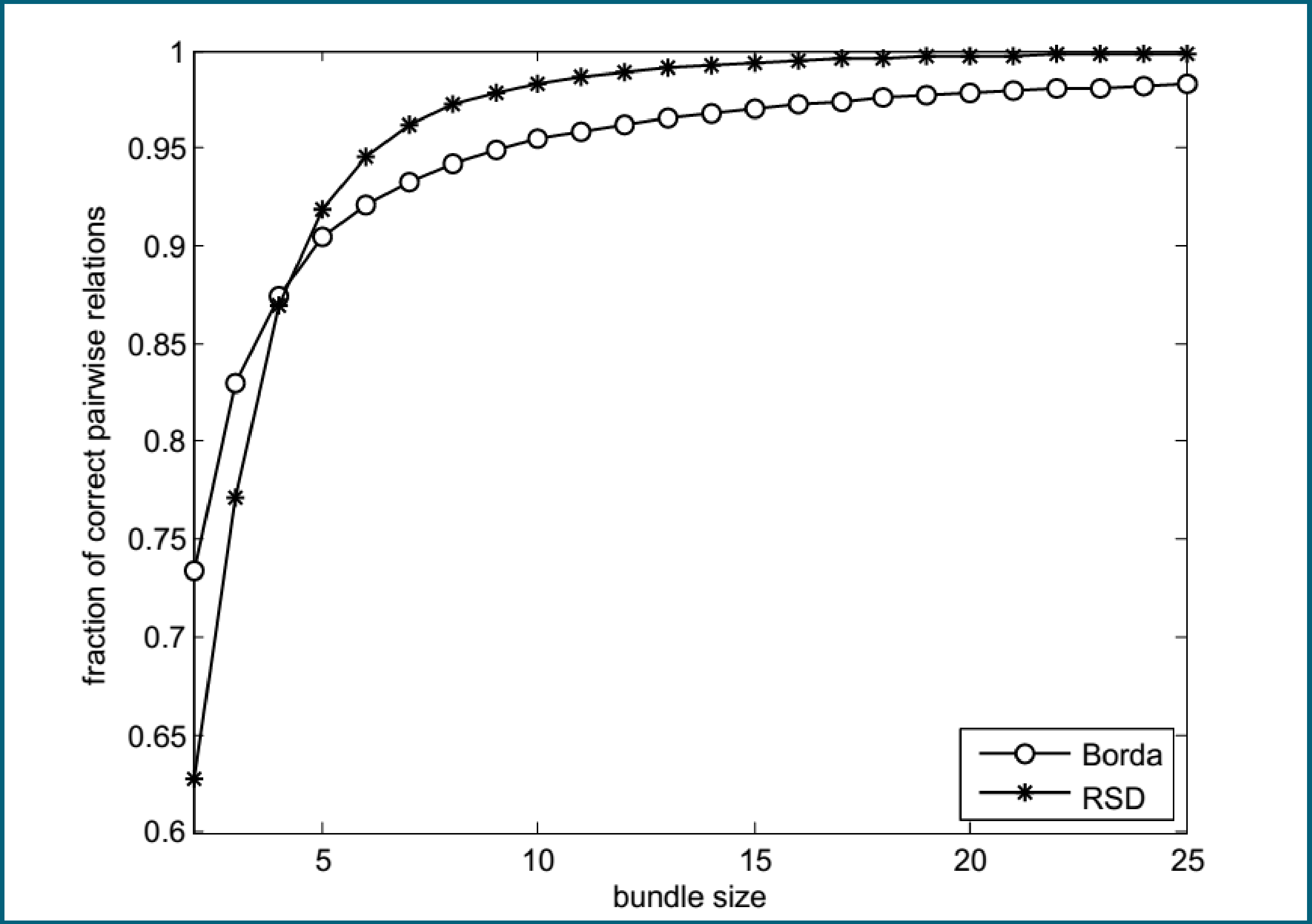
Borda *vs.* RSD (perfect grading)

Graph type		Random k-regular		Square – free		Copies of $K_{k,k}$	
k	n	Borda	RSD	Borda	RSD	Borda	RSD
2	1002	73.3	62.7	73.5	60.3	66.8	56.8
3	1001	83.0	77.2	83.2	66.0	73.1	60.2
4	1001	87.5	86.8	87.7	68.7	77.1	62.2
6	1023	92.0	94.6	92.1	72.7	81.6	65.2
8	1026	94.2	97.2	94.1	72.8	84.3	66.5
12	1064	96.3	98.9	96.6	76.0	87.3	68.5

Borda *vs.* RSD (perfect grading)

Graph type		Random k-regular		Square – free		Copies of $K_{k,k}$	
k	n	Borda	RSD	Borda	RSD	Borda	RSD
2	1002	73.3	62.7	73.5	60.3	66.8	56.8
3	1001	83.0	77.2	83.2	66.0	73.1	60.2
4	1001	87.5	86.8	87.7	68.7	77.1	62.2
6	1023	92.0	94.6	92.1	72.7	81.6	65.2
8	1026	94.2	97.2	94.1	72.8	84.3	66.5
12	1064	96.3	98.9	96.6	76.0	87.3	68.5

Borda *vs.* RSD (perfect grading)

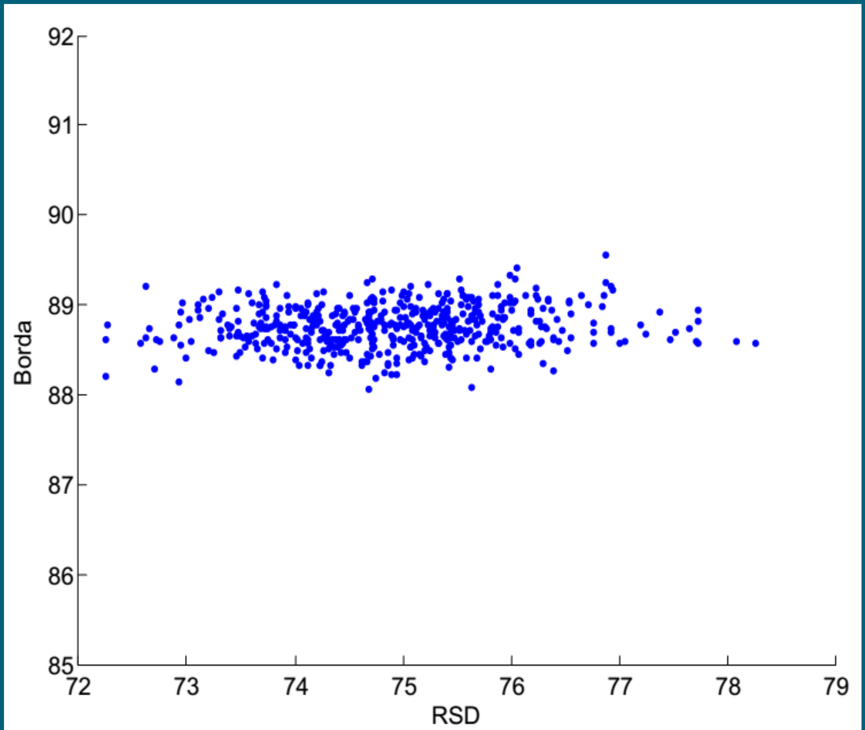


Producing noisy rankings (Mallows)

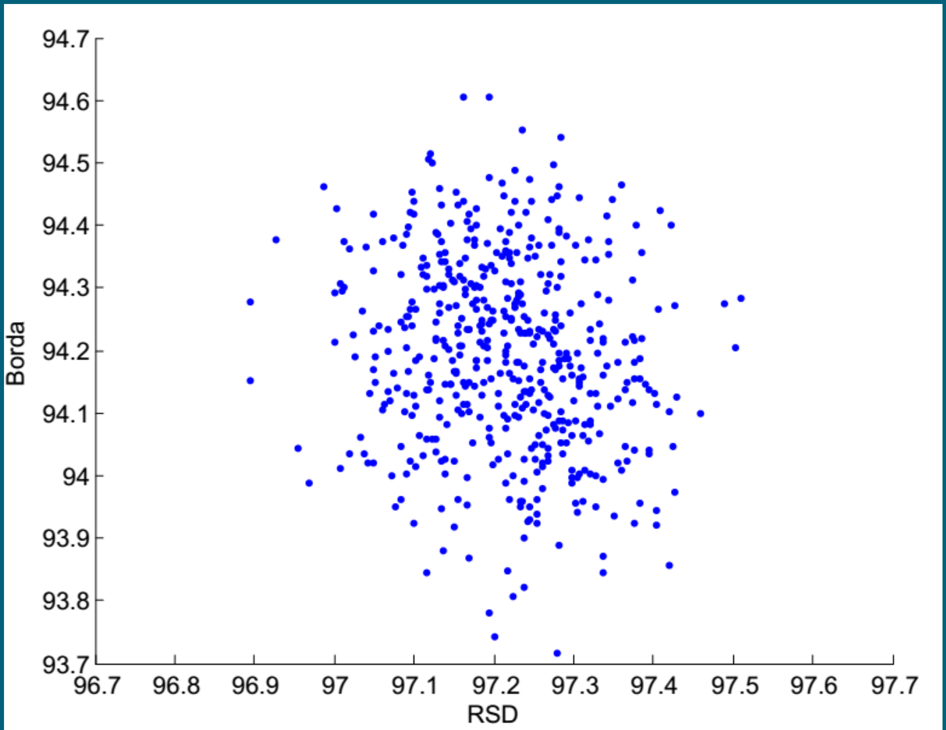
- Each student *i* has a quality q_i.
  - It affects his position in the ground truth and his ability to grade
- The ground truth $\succ$ is the ranking of the exam papers in decreasing order of quality
- Student *i* grades the exam papers in his bundle b_i according to the following procedure:
  - For every pair of exam papers (x, y) such that x $\succ$ y , set $x \succ_i y$ with probability q_i (and $y \succ_i x$ with probability 1-q_i)
  - If a cycle is created, then repeat from scratch

Imperfect grading

Noise of level 50%



Perfect Grading



Future research

- Analysis of Borda in the imperfect grading
- Analysis of RSD
- More realistic noise models
  - Random Utility Model
- Real world experiments
- Incentives (game the grading procedure)

More info

I. Caragiannis, G. A. Krimpas, and A. A. Voudouris. Aggregating partial rankings with applications to peer grading in massive online open courses. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 675–683, 2015.

More info

I. Caragiannis, G. A. Krimpas, and A. A. Voudouris. Aggregating partial rankings with applications to peer grading in massive online open courses. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 675–683, 2015.