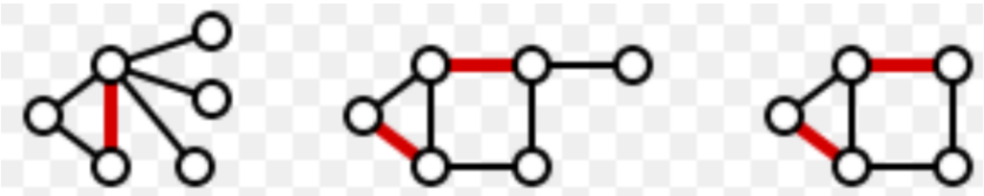
On Optimality and Scalability of Generalized Matching

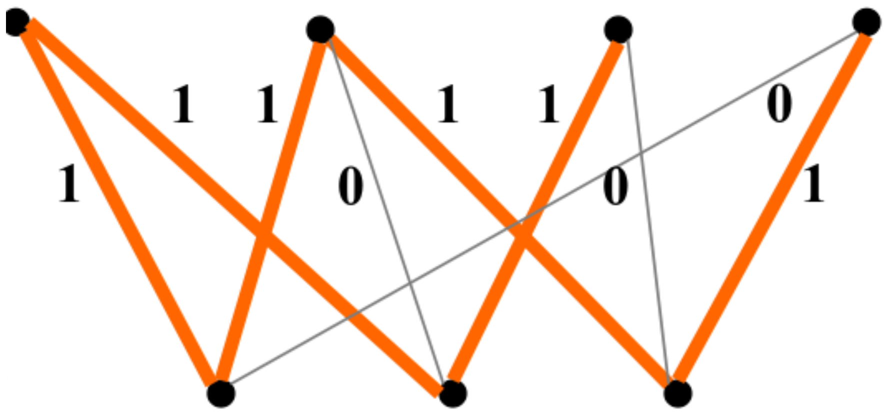
Cheng Chen

From Matching to "Generalized" Matching

- Matching: a matching M in a graph G is a set of non-adjacent edges (no two edges share a common vertex)

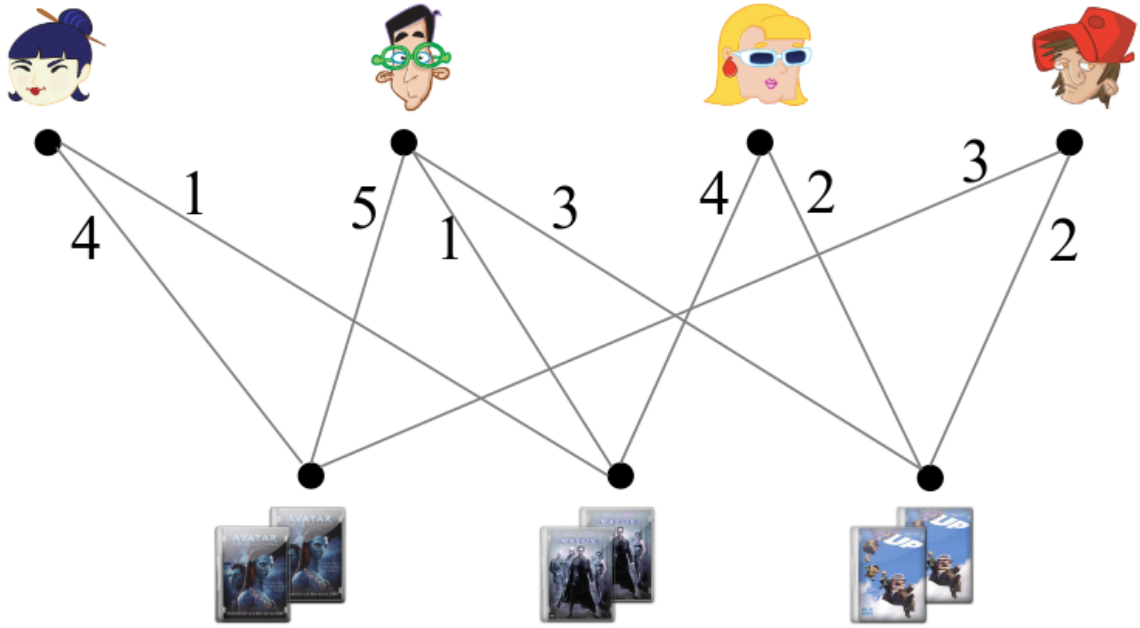


- Generalized Matching
  - Each node *v* is allowed to be connected to more than 1 neighbors
  - We study the bipartite graph



Bipartite Generalized Matching in Action (BGM)

- BGM has found applications in *Internet Advertising*, *Recommendation*, ...
- Assume we are running an online DVD rental website:
  - Each buyer/item has a **capacity**
  - Items are **in conflict** in terms of genre
  - Each buyer is recommended **at most** one item per genre (diversity)
  - Maximize the buyer utilities



Problem Definition

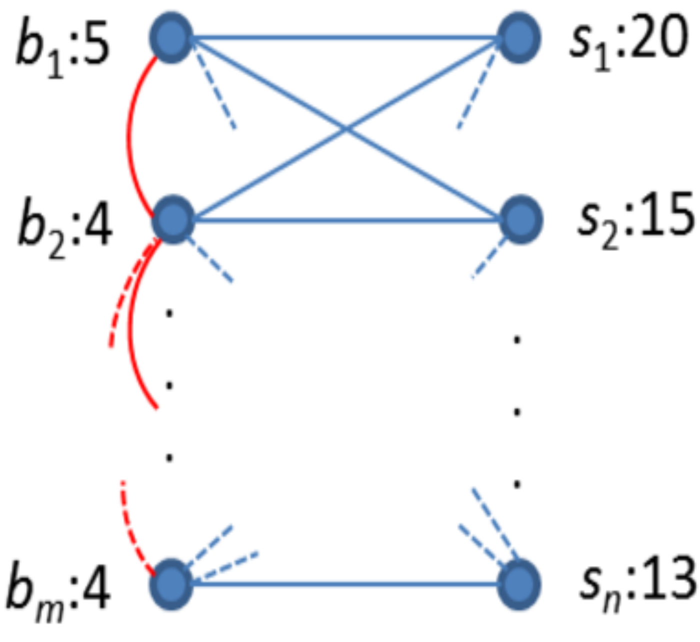
- Conflict-aware Constrained Recommendation (CAC-REC)
    - Input

1. An undirected, weighted graph $G = \langle (B, S), E \cup C, W \rangle$ with $E \subseteq B \times S$, $C \subseteq B \times B$ and weights $W : E \rightarrow \mathbb{R}^+$;
2. Degree constraints $D : B \cup S \rightarrow \mathbb{N}$;
3. A conflict threshold t.

- Objective

The goal in CAC-REC is to compute a maximum weight subgraph G' of G, $G' = ((B, S), E' \cup C', W)$, that satisfies the following two constraints:

1. For any i in $B \cup S$, $d_{G'}(i) \leq D(i)$.
2. For any k in S, $|\{(i, j) | (i, k), (j, k) \in E', (i, j) \in C'\}| \leq t$.



Problem Definition (Cont.)

$$\max_{X} \quad WX$$

$$s.t. \quad \mathbb{A}X(i) \le D(i), \forall i, 1 \le i \le m+n$$

$$x_{ij} \in \{0,1\}, \forall i,j, 1 \le i \le m, 1 \le j \le n,$$

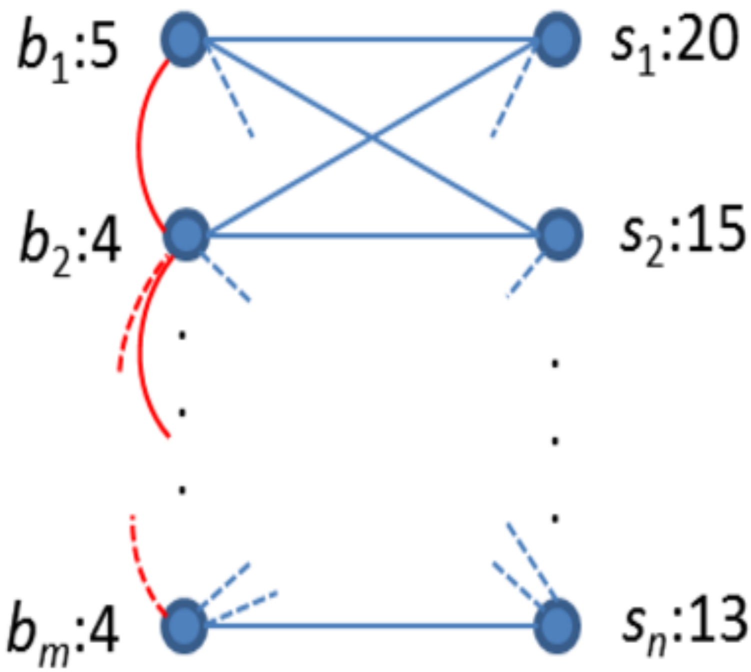
$$\sum_{(j_k, j_l) \in C} x_{ki} x_{li} \le t \; \forall \; i \in S$$

$$\max \quad w^\top x$$

$$\text{s.t.} \quad \boldsymbol{P}x \le p$$

$$\boldsymbol{C}x \ge c$$

$$x \ge 0,$$



CAC-REC is Hard

- Maximum Weight *b*-Matching can be solved in ***polynomial*** time via maximum-flow techniques.
  - For bipartite graph, the integral optimal solution can be found in polynomial-time

- Conflict constraints of CAC-REC, however, significantly increase the complexity, making it NP-hard.
  - There is a polynomial-time **reduction** from the NP-hard problem *Revenue Maximization in Interval Scheduling*

Algorithms: LP-based

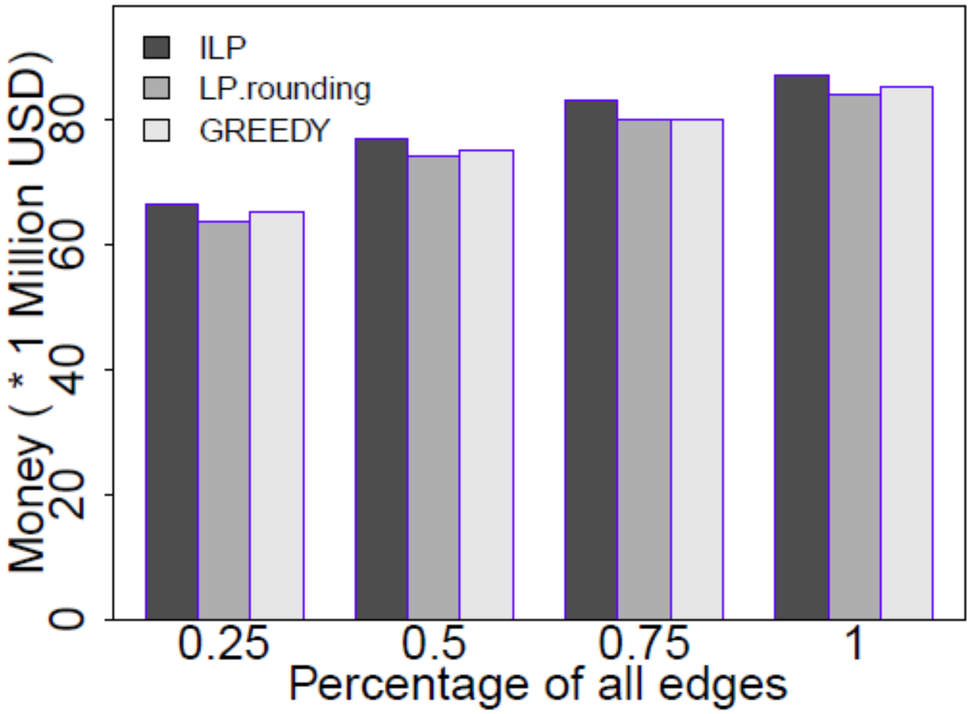
- Integer Linear Program (ILP)
    - Directly solved by ILP solver, such as Gurobi

- LP relaxation with rounding
    1. Solve the linear program relaxation to obtain optimal solution X (using Gurobi without integer constraints).
    2. Sort the fist mn elements of X from largest to smallest. Then round each non-zero value to 1 provided doing so does not violate the degree constraints or the conflict constraints. Otherwise, we set it to 0

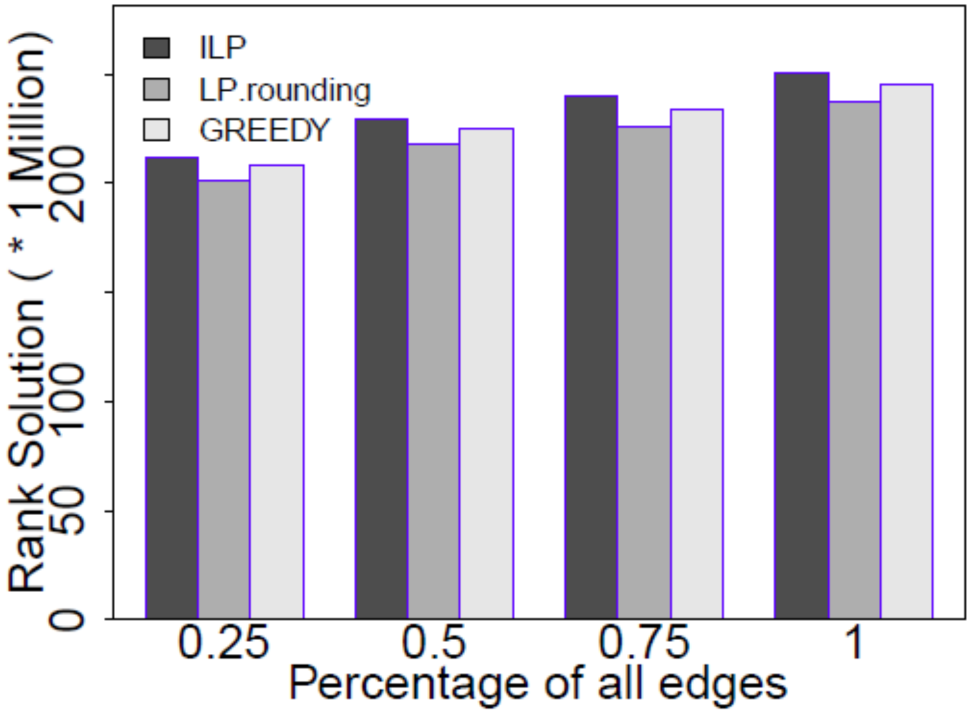
Algorithms: GREEDY

1. Sort all the edges in E by weights from largest to smallest
2. To construct the maximum weight subgraph G, consider every edge in the sorted list. Add this edge to G if doing so does not violate any degree constraint of conflict constraint.
3. Continue until we reach the end of the sorted list.

Experimental Results on LP-based Algorithms



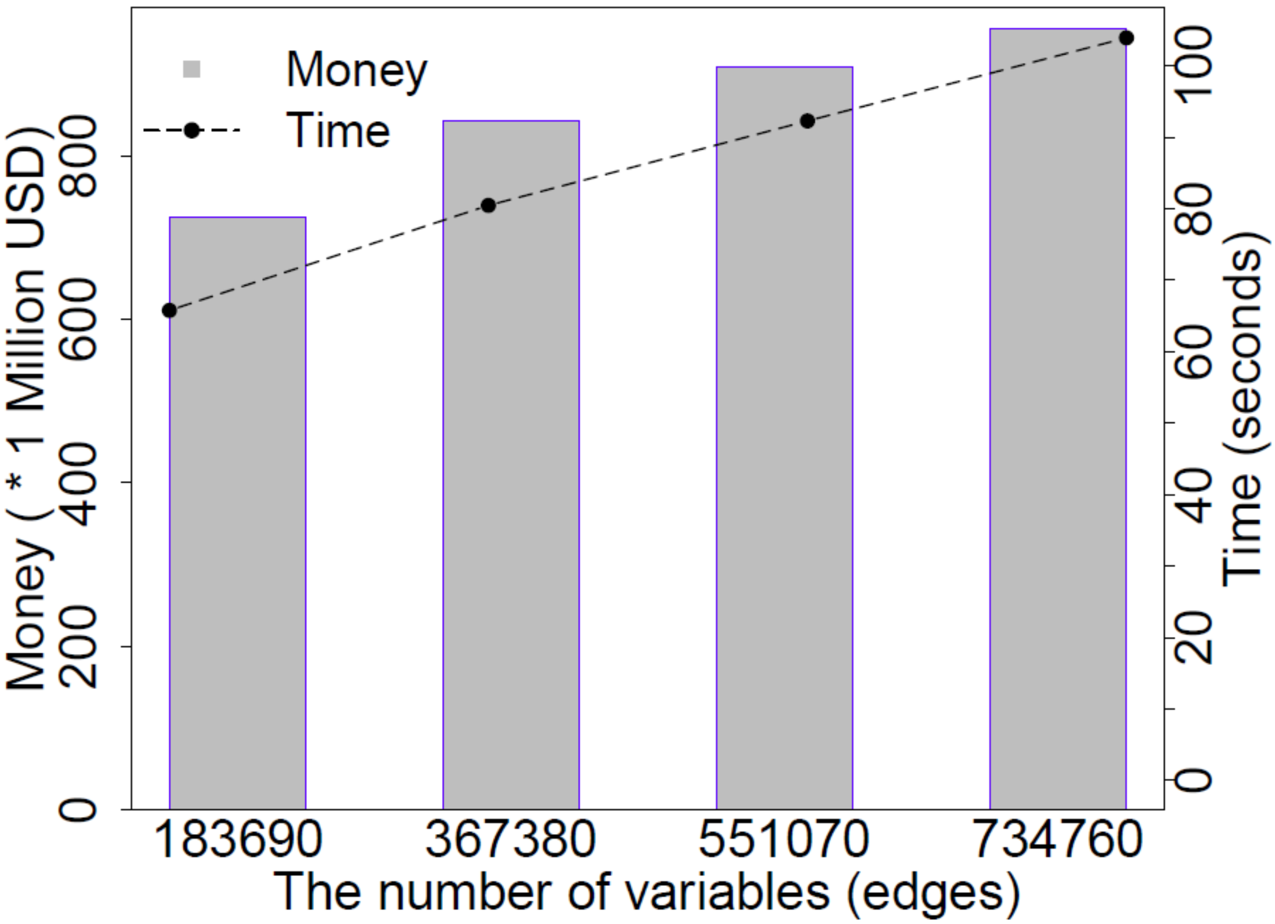
(a) Money Solution



(b) Rank Solution

Experimental Results on GREEDY

GREEDY is scalable to larger datasets.



Ongoing Work

- The need to cope with **BIG DATA**, when:
  - Gurobi fails
  - Global sorting fails
- More efficient LP solvers
- Parallel and Distributed Computing

Image Source: Forbes

References

- Iyer, Ganesh, David Soberman, and J. Miguel Villas-Boas. "The targeting of advertising." *Marketing Science* 24.3 (2005): 461-476.

- Armen S. Asratian, Tristan M. J. Denley, and Roland Häggkvist. 1998. *Bipartite Graphs and their Applications*. Cambridge University Press, New York, NY, USA.

- Aditya Parameswaran, Petros Venetis, and Hector Garcia-Molina. 2011. Recommendation systems with complex constraints: A course recommendation perspective. *ACM Trans. Inf. Syst.* 29, 4, Article 20 (December 2011), 33 pages.

- Julián Mestre. 2006. Greedy in approximation algorithms. In *Proceedings of the 14th conference on Annual European Symposium - Volume 14* (ESA'06), Yossi Azar and Thomas Erlebach (Eds.), Vol. 14. Springer-Verlag, London, UK, UK, 528-539.

- Maryam Karimzadehgan, ChengXiang Zhai, and Geneva Belford. 2008. Multi-aspect expertise matching for review assignment. In *Proceedings of the 17th ACM conference on Information and knowledge management* (CIKM '08). ACM, New York, NY, USA, 1113-1122.