

Étude combinatoire et algorithmique de médianes de permutations sous la distance de Kendall-Tau

Robin Milosz

Département d'Informatique et de Recherche Opérationnelle (DIRO)
Université de Montréal

29 août 2017



Plan

Introduction

Définition du problème

Motivation

Travaux antérieurs

Travail publié

- 1) Heuristique
- 2) Réduction d'espace
- 3) Branch-and-Bound

Résultats préliminaires et travaux futurs

- 1) Programmation en nombres entiers
- 2) Formule d'approximation
- 3) Heuristique "MedianGame"
- 4) Grandes instances
- 5) Classements avec égalité

Conclusion et calendrier





Définition: Le problème de la médiane de permutations

Étant donné un ensemble de permutations on veut trouver une permutation qui est la "plus proche" de cet ensemble, selon une distance donnée.



Définition: Distance de Kendall- τ

La **distance de Kendall- τ** ¹, notée d_{KT} , compte le nombre de paires en désaccord entre deux permutations.

$$d_{KT}(\pi, \sigma) = \#\{(i, j) | i < j \text{ and } [(\pi[i] < \pi[j] \text{ and } \sigma[i] > \sigma[j]) \\ \text{or } (\pi[i] > \pi[j] \text{ and } \sigma[i] < \sigma[j])]\},$$

$$d_{KT}(\pi, \mathcal{A}) = \sum_{\sigma \in \mathcal{A}} d_{KT}(\pi, \sigma).$$

¹M. Kendall, *A New Measure of Rank Correlation*, Biometrika, 30, pp.81-89, 1938.

Définition: Distance de Kendall- τ

La **distance de Kendall- τ** ¹, notée d_{KT} , compte le nombre de paires en désaccord entre deux permutations.

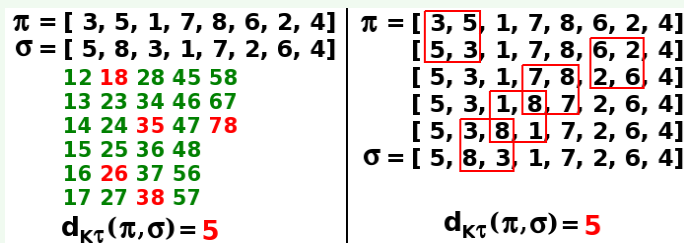


Figure: La distance de Kendall- τ : exemple sur deux permutations de $\{1, 2, \dots, n\}$

¹M. Kendall, *A New Measure of Rank Correlation*, Biometrika, 30, pp.81-89, 1938.

Définition: Médiane sous la distance de Kendall- τ ²

Étant donné $\mathcal{A} \subseteq \mathcal{S}_n$, on veut trouver une permutation π^* de \mathcal{S}_n telle que

$$d_{KT}(\pi^*, \mathcal{A}) \leq d_{KT}(\pi, \mathcal{A}), \quad \forall \pi \in \mathcal{S}_n.$$

Notation récurrentes:

$$\#A = m$$

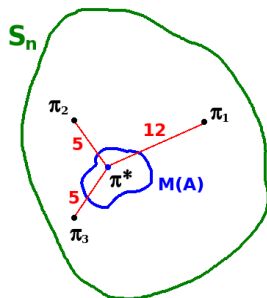
permutations de taille n

$M(A)$: ensemble des permutations médianes

²J. Kemeny, *Mathematics without numbers*, Daedalus 88 , pp. 577-591, 1959.

Définition: Médiane sous la distance de Kendall- τ ²

$$A = \{ \pi_1, \pi_2, \pi_3 \} \quad \pi^* \in M(A)$$



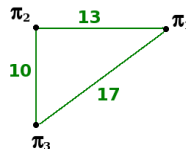
$$d_{\text{K}\tau}(\pi^*, A) = 12 + 5 + 5 = 22$$

$$\pi_1 = [7, 8, 2, 3, 6, 1, 5, 4]$$

$$\pi_2 = [3, 5, 1, 7, 8, 6, 2, 4]$$

$$\pi_3 = [5, 8, 3, 4, 1, 2, 7, 6]$$

$$\pi^* = [5, 8, 3, 1, 7, 2, 6, 4]$$



²J. Kemeny, *Mathematics without numbers*, Daedalus 88 , pp. 577-591, 1959.

Motivations et applications

Applications de la médiane de permutations:

- Bio-informatique: méta-recherche, construction de cartes génétiques ³
- Classements: sports, choix sociaux/politiques, moteurs de recherche et films ⁴

³B. N. Jackson, P. S. Schnable, and S. Aluru, *Consensus genetic maps as median orders from inconsistent sources*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 5(2), pp.161-171, 2008.

⁴C. Dwork, R. Kumar, M. Naor and D. Sivakumar, *Rank Aggregation Methods for the Web*, in proceedings of the 10th WWW, pp.613-622, 2001.

Motivations et applications

Applications de la médiane de permutations:

- Bio-informatique: méta-recherche, construction de cartes génétiques ³
- Classements: sports, choix sociaux/politiques, moteurs de recherche et films ⁴

³B. N. Jackson, P. S. Schnable, and S. Aluru, *Consensus genetic maps as median orders from inconsistent sources*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 5(2), pp.161-171, 2008.


⁴C. Dwork, R. Kumar, M. Naor and D. Sivakumar, *Rank Aggregation Methods for the Web*, in proceedings of the 10th WWW, pp.613-622, 2001.

Motivations et applications

Applications de la médiane de permutations:

- Bio-informatique: méta-recherche, construction de cartes génétiques ³
- Classements: sports, choix sociaux/politiques, moteurs de recherche et films ⁴

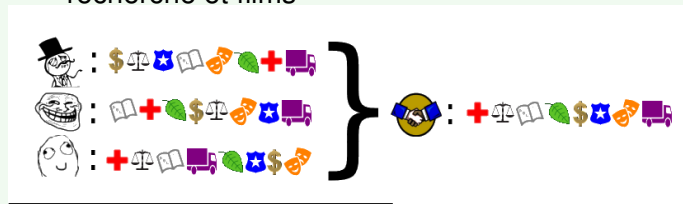
³B. N. Jackson, P. S. Schnable, and S. Aluru, *Consensus genetic maps as median orders from inconsistent sources*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 5(2), pp.161-171, 2008.

⁴C. Dwork, R. Kumar, M. Naor and D. Sivakumar, *Rank Aggregation Methods for the Web*, in proceedings of the 10th WWW, pp.613-622, 2001. 

Motivations et applications

Applications de la médiane de permutations:

- Bio-informatique: méta-recherche, construction de cartes génétiques ³
- Classements: sports, choix sociaux/politiques, moteurs de recherche et films ⁴



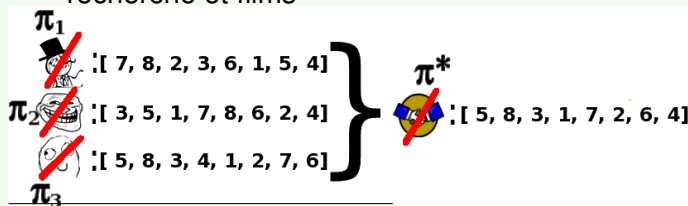
³B. N. Jackson, P. S. Schnable, and S. Aluru, *Consensus genetic maps as median orders from inconsistent sources*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 5(2), pp.161-171, 2008.

⁴C. Dwork, R. Kumar, M. Naor and D. Sivakumar, *Rank Aggregation Methods for the Web*, in proceedings of the 10th WWW, pp.613-622, 2001.

Motivations et applications

Applications de la médiane de permutations:

- Bio-informatique: méta-recherche, construction de cartes génétiques ³
- Classements: sports, choix sociaux/politiques, moteurs de recherche et films ⁴



³B. N. Jackson, P. S. Schnable, and S. Aluru, *Consensus genetic maps as median orders from inconsistent sources*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 5(2), pp.161-171, 2008.

⁴C. Dwork, R. Kumar, M. Naor and D. Sivakumar, *Rank Aggregation Methods for the Web*, in proceedings of the 10th WWW, pp.613-622, 2001.

Travaux antérieurs pertinents

- Kendall (1938): distance
- Kemeny (1959): problème
- Young (78-95): caractéristiques
- Bartholdi (1989): NP-Difficile
- Dwork et al. (2001): NP-Difficile $m \geq 4$, heuristique MC
- Niedermeier, Betzler, Simjour, Karpinski, Schudy, Nishimura et al. (2008-2014): algorithmes à paramètres fixes (dist. médiane, dist. max A, dist. moy. A, écart pos. max, dist. médiane/m)

Travaux antérieurs pertinents

- Kendall (1938): distance
- Kemeny (1959): problème
- Young (78-95): caractéristiques
- Bartholdi (1989): NP-Difficile
- Dwork et al. (2001): NP-Difficile $m \geq 4$, heuristique MC
- Niedermeier, Betzler, Simjour, Karpinski, Schudy, Nishimura et al. (2008-2014): algorithmes à paramètres fixes (dist. médiane, dist. max A, dist. moy. A, écart pos. max, dist. médiane/m)

Travaux antérieurs pertinents

- Kendall (1938): distance
- Kemeny (1959): problème
- Young (78-95): caractéristiques
- Bartholdi (1989): NP-Difficile
- Dwork et al. (2001): NP-Difficile $m \geq 4$, heuristique MC
- Niedermeier, Betzler, Simjour, Karpinski, Schudy, Nishimura et al. (2008-2014): algorithmes à paramètres fixes (dist. médiane, dist. max A, dist. moy. A, écart pos. max, dist. médiane/m)

Travaux antérieurs pertinents

- Kendall (1938): distance
- Kemeny (1959): problème
- Young (78-95): caractéristiques
- Bartholdi (1989): NP-Difficile
- Dwork et al. (2001): NP-Difficile $m \geq 4$, heuristique MC
- Niedermeier, Betzler, Simjour, Karpinski, Schudy, Nishimura et al. (2008-2014): algorithmes à paramètres fixes (dist. médiane, dist. max A, dist. moy. A, écart pos. max, dist. médiane/m)

Travaux antérieurs pertinents

- Kendall (1938): distance
- Kemeny (1959): problème
- Young (78-95): caractéristiques
- Bartholdi (1989): NP-Difficile
- Dwork et al. (2001): NP-Difficile $m \geq 4$, heuristique MC
- Niedermeier, Betzler, Simjour, Karpinski, Schudy, Nishimura et al. (2008-2014): algorithmes à paramètres fixes (dist. médiane, dist. max A, dist. moy. A, écart pos. max, dist. médiane/m)

Travaux antérieurs pertinents

- Kendall (1938): distance
- Kemeny (1959): problème
- Young (78-95): caractéristiques
- Bartholdi (1989): NP-Difficile
- Dwork et al. (2001): NP-Difficile $m \geq 4$, heuristique MC
- Niedermeier, Betzler, Simjour, Karpinski, Schudy, Nishimura et al. (2008-2014): algorithmes à paramètres fixes (dist. médiane, dist. max A, dist. moy. A, écart pos. max, dist. médiane/m)

Travaux antérieurs pertinents

- Conitzer, Davenport et Kalagnanam (2006): borne inférieure, heuristique et solveurs
- van Zuylen et Schalekamp (2009): heuristiques de tri
- Blin et al. (2011): contraintes et heuristique
- Meilă et al. (2012): Travail de comparaison
- Betzler et al. (2014): Réduction d'espace
- Milosz et Hamel (2016): Réduction d'espace

Travaux antérieurs pertinents

- Conitzer, Davenport et Kalagnanam (2006): borne inférieure, heuristique et solveurs
- van Zuylen et Schalekamp (2009): heuristiques de tri
- Blin et al. (2011): contraintes et heuristique
- Meilă et al. (2012): Travail de comparaison
- Betzler et al. (2014): Réduction d'espace
- Milosz et Hamel (2016): Réduction d'espace

Travaux antérieurs pertinents

- Conitzer, Davenport et Kalagnanam (2006): borne inférieure, heuristique et solveurs
- van Zuylen et Schalekamp (2009): heuristiques de tri
- Blin et al. (2011): contraintes et heuristique
- Meilă et al. (2012): Travail de comparaison
- Betzler et al. (2014): Réduction d'espace
- Milosz et Hamel (2016): Réduction d'espace

Travaux antérieurs pertinents

- Conitzer, Davenport et Kalagnanam (2006): borne inférieure, heuristique et solveurs
- van Zuylen et Schalekamp (2009): heuristiques de tri
- Blin et al. (2011): contraintes et heuristique
- Meilă et al. (2012): Travail de comparaison
- Betzler et al. (2014): Réduction d'espace
- Milosz et Hamel (2016): Réduction d'espace

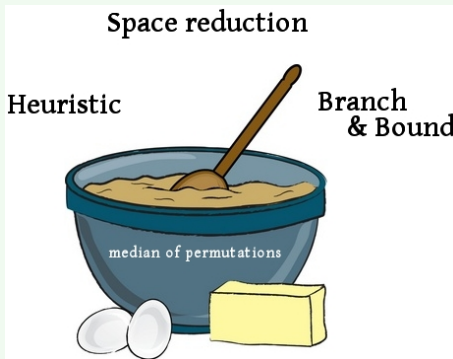
Travaux antérieurs pertinents

- Conitzer, Davenport et Kalagnanam (2006): borne inférieure, heuristique et solveurs
- van Zuylen et Schalekamp (2009): heuristiques de tri
- Blin et al. (2011): contraintes et heuristique
- Meilă et al. (2012): Travail de comparaison
- Betzler et al. (2014): Réduction d'espace
- Milosz et Hamel (2016): Réduction d'espace

Travaux antérieurs pertinents

- Conitzer, Davenport et Kalagnanam (2006): borne inférieure, heuristique et solveurs
- van Zuylen et Schalekamp (2009): heuristiques de tri
- Blin et al. (2011): contraintes et heuristique
- Meilă et al. (2012): Travail de comparaison
- Betzler et al. (2014): Réduction d'espace
- Milosz et Hamel (2016): Réduction d'espace

Notre plan sur 3 approches



"Heuristic, Branch-and-Bound Solver and Improved Space Reduction for the Median of Permutations Problem"
IWOCA2017

1) Heuristique

Viser à trouver la vraie médiane (ou proche de) pour des instances du problème de petite et moyenne taille.

1) Heuristique: Recuit Simulé (Simulated Annealing)

Pseudo-code

currentSol \leftarrow initialSol

for $i \leftarrow 1 \dots nbSteps$

 alternativeSol \leftarrow neighbour(currentSol)

$\Delta E \leftarrow d_{KT}(\text{alternativeSol}, \mathcal{A}) - d_{KT}(\text{currentSol}, \mathcal{A})$

 if ($\Delta E \leq 0$) "meilleure solution"

 currentSol \leftarrow alternativeSol

 else if ($rand[0, 1] < e^{-\Delta E/T}$) "pire solution"

 currentSol \leftarrow alternativeSol

$T \leftarrow T \times \alpha$

return currentSol

1) Heuristique: Recuit Simulé - exemple

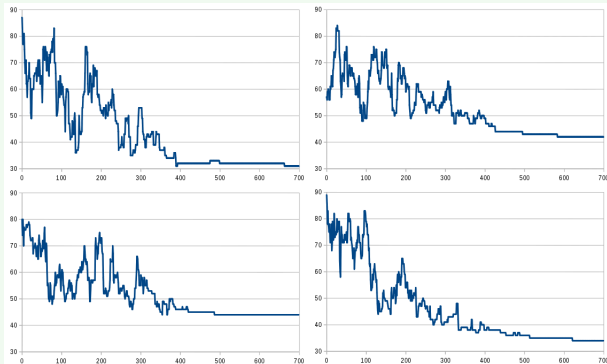


Figure: SA: Distance de Kendall- τ de la solution courante au fil des itérations.

1) Heuristique: Résultats

- testage intensif pour trouver les paramètres (plus de 407 heures sur 14 processeurs Intel i7 @ 2.93GHz)
- vraie médiane "presque" tout le temps pour $n \leq 38$ et $m \leq 50$ (96% et +)

2) Réduction d'espace

Améliorer la réduction d'espace de recherche en combinant
plusieurs méthodes

première méthode: contraintes MOT



2) Réduction d'espace: Major Order Theorems⁵ (Théorèmes de l'Ordre Majeur)

$A = \{ [7, \textcircled{8}, \textcircled{2}, 3, 6, 1, 5, 4],$
 $[3, 5, 1, 7, \textcircled{8}, 6, \textcircled{2}, 4],$
 $[5, \textcircled{8}, 3, 4, 1, \textcircled{2}, 7, 6] \}$

$\textcircled{8} < \textcircled{2}$

always

3 < 1

3 < 4

3 < 6

5 < 4

7 < 6

8 < 2

8 < 4

8 < 6

3 < 1

3 < 4

3 < 6

5 < 4

7 < 6

8 < 2

8 < 4

8 < 6

3 < 1

3 < 4

3 < 6

5 < 4

7 < 6

8 < 2

8 < 4

8 < 6

3 < 1

3 < 4

3 < 6

5 < 4

7 < 6

8 < 2

8 < 4

8 < 6

3 < 1

3 < 4

3 < 6

5 < 4

7 < 6

8 < 2

8 < 4

8 < 6

3 < 1

3 < 4

3 < 6

5 < 4

7 < 6

8 < 2

8 < 4

8 < 6

3 < 1

3 < 4

3 < 6

5 < 4

7 < 6

8 < 2

8 < 4

8 < 6

⁵R. Milosz and S. Hamel, *Medians of Permutations: Building Constraints*, LNCS 9602, Second International Conference on Algorithms and Discrete Applied Mathematics, pp. 264–276, 2016.

2) Réduction d'espace: Major Order Theorems⁵ (Théorèmes de l'Ordre Majeur)

$A = \{ [7, 8, 2, 3, 6, \textcircled{1}, 5, \textcircled{4}],$
 $[3, 5, \textcircled{1}, 7, 8, 6, 2, \textcircled{4}],$
 $[5, 8, 3, \textcircled{4}, \textcircled{1}, 2, 7, 6] \}$

$\textcircled{1} < \textcircled{4}$

always	MOT 1.0
3<1	1<4
3<4	1<6
3<6	2<6
5<4	3<2
7<6	5<1
8<2	7<2
8<4	
8<6	add (closure)
	5<6

⁵R. Milosz and S. Hamel, *Medians of Permutations: Building Constraints*, LNCS 9602, Second International Conference on Algorithms and Discrete Applied Mathematics, pp. 264–276, 2016.

2) Réduction d'espace: Major Order Theorems⁵ (Théorèmes de l'Ordre Majeur)

$A = \{ [7, 8, \textcircled{2}, \textcolor{red}{3}, \textcolor{red}{6}, \textcolor{red}{1}, \textcircled{5}, 4],$
 $[3, \textcircled{5}, \textcolor{red}{1}, 7, 8, \textcolor{red}{6}, \textcircled{2}, 4],$
 $[\textcircled{5}, 8, \textcolor{red}{3}, 4, 1, \textcircled{2}, 7, 6] \}$

$\textcircled{5} < \textcircled{2}$

always	MOT 1.0	MOT 2.0
3<1	1<4	2<4
3<4	1<6	5<2
3<6	2<6	7<4
5<4	3<2	
7<6	5<1	
8<2	7<2	
8<4		
8<6	add (closure)	
	5<6	

⁵R. Milosz and S. Hamel, *Medians of Permutations: Building Constraints*, LNCS 9602, Second International Conference on Algorithms and Discrete Applied Mathematics, pp. 264–276, 2016.



2) Réduction d'espace: Major Order Theorems⁵ (Théorèmes de l'Ordre Majeur)

$A = \{ [7, 8, \textcircled{2}, \textcolor{red}{3}, \textcolor{red}{6}, \textcircled{1}, 5, 4],$
 $[3, 5, \textcircled{1}, \textcolor{green}{7}, \textcolor{green}{8}, \textcolor{red}{6}, \textcircled{2}, 4],$
 $[5, 8, 3, 4, \textcircled{1}, \textcircled{2}, 7, 6] \}$

$\textcircled{1} < \textcircled{2}$

always	MOT 1.0	MOT 2.0	MOT 3.0
→ 3 < 1	1 < 4	2 < 4	1 < 2
3 < 4	1 < 6	5 < 2	5 < 7
3 < 6	2 < 6	7 < 4	6 < 4
5 < 4	→ 3 < 2		
7 < 6	5 < 1		
8 < 2	7 < 2		
8 < 4			
8 < 6	add (closure)		
	5 < 6		

⁵R. Milosz and S. Hamel, *Medians of Permutations: Building Constraints*, LNCS 9602, Second International Conference on Algorithms and Discrete Applied Mathematics, pp. 264–276, 2016.



2) Réduction d'espace

Améliorer la réduction d'espace de recherche en combinant
plusieurs méthodes

deuxième méthode: borne inférieure



2) Réduction d'espace: borne inférieure de Conitzer et al. ⁶

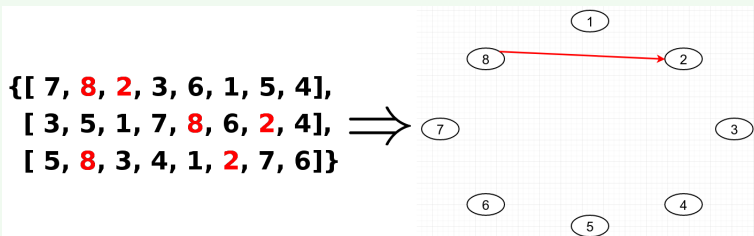


Figure: Graphe de tournoi: construction

⁶V. Conitzer, A. Davenport, and J. Kalagnanam, *Improved bounds for computing Kemeny rankings*, Proceedings of the 21st conference on Artificial intelligence - Volume 1, AAAI'06, pp. 620–626, 2006.

2) Réduction d'espace: borne inférieure de Conitzer et al. ⁶

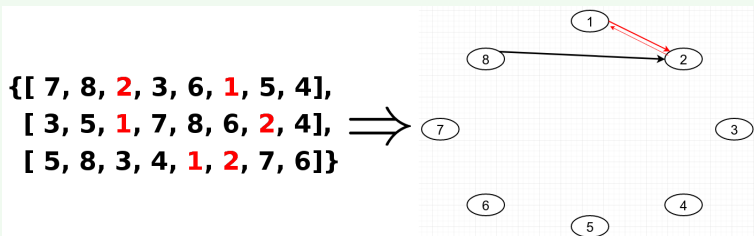


Figure: Graphe de tournoi: construction

⁶V. Conitzer, A. Davenport, and J. Kalagnanam, *Improved bounds for computing Kemeny rankings*, Proceedings of the 21st conference on Artificial intelligence - Volume 1, AAAI'06, pp. 620–626, 2006.

2) Réduction d'espace: borne inférieure de Conitzer et al. ⁶

{ [7, 8, 2, 3, 6, 1, 5, 4],
[3, 5, 1, 7, 8, 6, 2, 4],
[5, 8, 3, 4, 1, 2, 7, 6] }

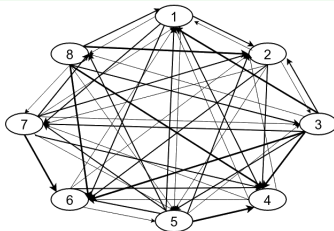


Figure: Graphe de tournoi: borne inférieure = 0

⁶V. Conitzer, A. Davenport, and J. Kalagnanam, *Improved bounds for computing Kemeny rankings*, Proceedings of the 21st conference on Artificial intelligence - Volume 1, AAAI'06, pp. 620–626, 2006.

2) Réduction d'espace: borne inférieure de Conitzer et al. ⁶

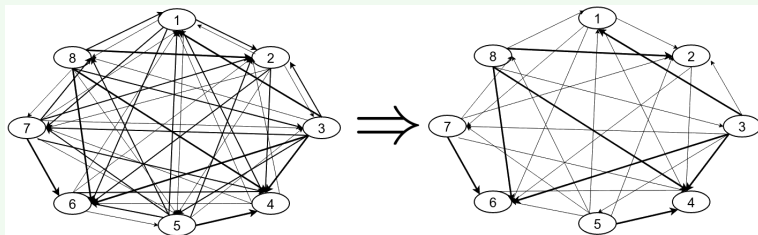


Figure: Graphe de tournoi: borne inférieure = 20

⁶V. Conitzer, A. Davenport, and J. Kalagnanam, *Improved bounds for computing Kemeny rankings*, Proceedings of the 21st conference on Artificial intelligence - Volume 1, AAAI'06, pp. 620–626, 2006.

2) Réduction d'espace: borne inférieure de Conitzer et al. ⁶

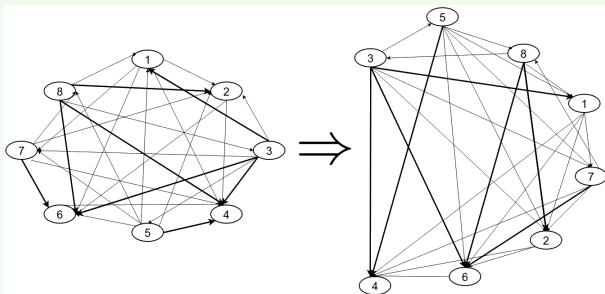


Figure: Graphe de tournoi: borne inférieure = 20

⁶V. Conitzer, A. Davenport, and J. Kalagnanam, *Improved bounds for computing Kemeny rankings*, Proceedings of the 21st conference on Artificial intelligence - Volume 1, AAAI'06, pp. 620–626, 2006.

2) Réduction d'espace: borne inférieure de Conitzer et al.⁶

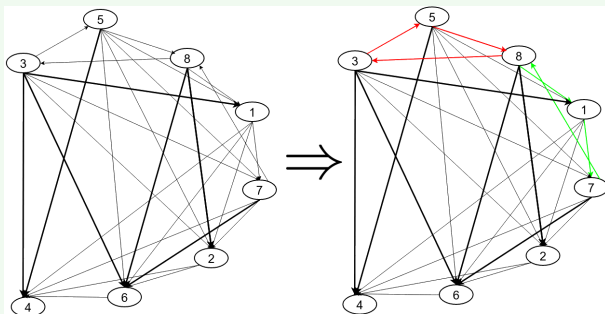


Figure: Graphe de tournoi: borne inférieure = 20 + 1 + 1

⁶V. Conitzer, A. Davenport, and J. Kalagnanam, *Improved bounds for computing Kemeny rankings*, Proceedings of the 21st conference on Artificial intelligence - Volume 1, AAAI'06, pp. 620–626, 2006.

2) Réduction d'espace: borne inférieure de Conitzer et al. ⁶

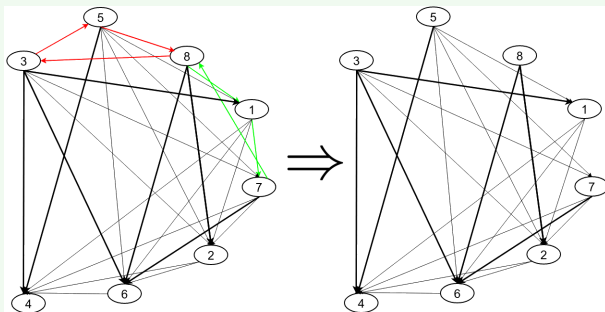


Figure: Graphe de tournoi: borne inférieure = 22

⁶V. Conitzer, A. Davenport, and J. Kalagnanam, *Improved bounds for computing Kemeny rankings*, Proceedings of the 21st conference on Artificial intelligence - Volume 1, AAAI'06, pp. 620–626, 2006.

2) Réduction d'espace

Améliorer la réduction d'espace de recherche en combinant
plusieurs méthodes

combinaison des 2 premières méthodes avec une approximation

2) Réduction d'espace: LUBC

Lower Upper Bound Constraints (Contraintes de borne inférieure et supérieure)

- Utiliser les contraintes MOT pour améliorer la borne inférieure de Conitzer et al.
- Supposer des nouvelles contraintes et utiliser la preuve par contradiction

$$i \prec j ? \quad C' = C \cup \{j \prec i\}$$

$$\text{borneInf}(C') > \text{approx. ?} \quad \begin{cases} \text{si oui} & i \prec j \text{ vérifiée} \\ \text{si non} & \text{rien} \end{cases}$$

- Répéter (MOT & LUBC)

2) Réduction d'espace: LUBC

Lower Upper Bound Constraints (Contraintes de borne inférieure et supérieure)

- Utiliser les contraintes MOT pour améliorer la borne inférieure de Conitzer et al.
- Supposer des nouvelles contraintes et utiliser la preuve par contradiction

$$i \prec j ? \quad C' = C \cup \{j \prec i\}$$

$$\text{borneInf}(C') > \text{approx. ?} \quad \begin{cases} \text{si oui} & i \prec j \text{ vérifiée} \\ \text{si non} & \text{rien} \end{cases}$$

- Répéter (MOT & LUBC)

2) Réduction d'espace: LUBC

Lower Upper Bound Constraints (Contraintes de borne inférieure et supérieure)

- Utiliser les contraintes MOT pour améliorer la borne inférieure de Conitzer et al.
- Supposer des nouvelles contraintes et utiliser la preuve par contradiction

$$i \prec j ? \quad C' = C \cup \{j \prec i\}$$

$$\text{borneInf}(C') > \text{approx. ?} \quad \begin{cases} \text{si oui} & i \prec j \text{ vérifiée} \\ \text{si non} & \text{rien} \end{cases}$$

- Répéter (MOT & LUBC)



2) Réduction d'espace: Résultats

m	n = 15				n = 30				n = 45			
	MOT	MOT+LUBC	MOTe	MOTe+LUBC	MOT	MOT+LUBC	MOTe	MOTe+LUBC	MOT	MOT+LUBC	MOTe	MOTe+LUBC
3	0.635	0.921	0.878	0.966	0.506	0.725	0.702	0.873	0.447	0.513	0.604	0.669
5	0.595	0.913	0.800	0.954	0.444	0.605	0.595	0.766	0.361	0.381	0.490	0.516
10	0.524	0.845	0.824	0.951	0.384	0.616	0.549	0.747	0.310	0.347	0.419	0.454
15	0.579	0.939	0.704	0.953	0.400	0.569	0.490	0.668	0.316	0.328	0.383	0.398
20	0.540	0.884	0.747	0.945	0.383	0.597	0.493	0.697	0.306	0.327	0.377	0.399
25	0.581	0.949	0.680	0.923	0.398	0.579	0.465	0.653	0.312	0.322	0.363	0.375
30	0.550	0.904	0.720	0.947	0.386	0.589	0.472	0.674	0.301	0.317	0.362	0.379
35	0.584	0.955	0.668	0.961	0.396	0.582	0.453	0.642	0.309	0.320	0.345	0.356
40	0.556	0.915	0.702	0.950	0.388	0.590	0.461	0.660	0.307	0.323	0.354	0.371
45	0.587	0.959	0.660	0.964	0.397	0.586	0.448	0.638	0.306	0.317	0.350	0.363
50	0.562	0.923	0.691	0.952	0.388	0.589	0.456	0.654	0.301	0.314	0.352	0.366

Table: Efficacité des méthodes MOT, MOT+LUBC, MOTe, MOTe+LUBC en termes de proportion de paires d'éléments ordonnées, sur des ensembles de permutations aléatoires.



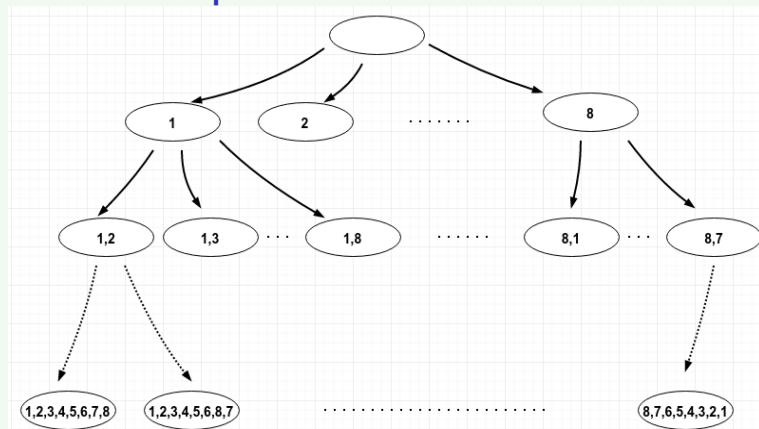
3) Branch-and-Bound

Résoudre le problème avec un algorithme exact



3) Branch-and-Bound

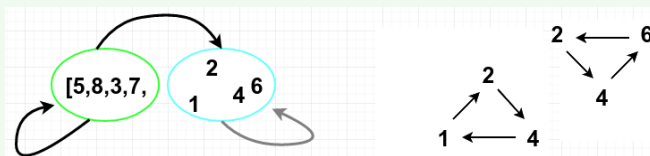
Construire les permutations



3) Branch-and-Bound

Borne inférieure

- borne inférieure de base
- utilisant des 3-cycles précalculés



3) Branch-and-Bound

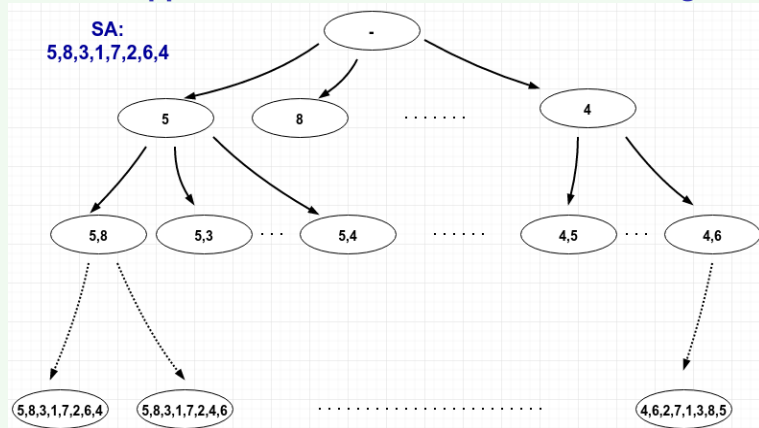
Élagage

- les triplets interdits
- contraintes MOT et LUBC
- "TopScores": structure de données pour éviter les répétitions

<u>Triplets</u>	<u>Constraints</u>	<u>TopScores</u>
1 2 3 x	3<1 3<2	<00010110,13>
1 2 4 ○	3<4 5<1	<00101011,15>
1 2 5 x	3<6 7<2	<00110100,11>
1 2 6 ○	5<4 5<6	<01000100, 8>
1 2 7 x	7<6 2<4	<01001110,16>
1 2 8 x	8<2 5<2	<01010001, 9>
1 3 2 x	8<4 7<4	<10001100,11>
...	8<6 1<2	<11000010,10>
8 7 5 x	1<4 5<7	<11100000,13>
8 7 6 x	1<6 6<4	
	2<6	

3) Branch-and-Bound

Utiliser l'approximation du recuit simulé comme guide





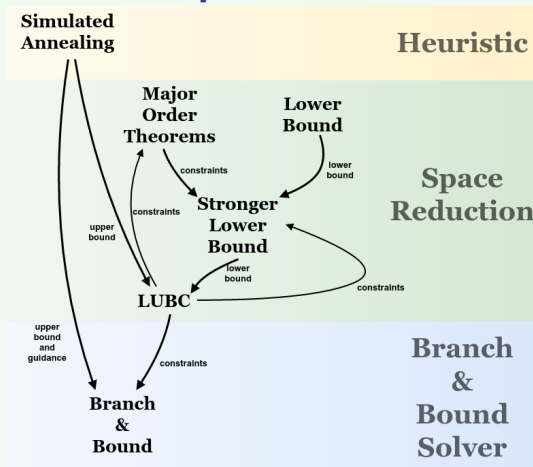
3) Branch-and-Bound: Résultats

- résout en temps raisonnable (secondes) des instances de m permutations de n pour $n \leq 38$ et $m \leq 50$
- comparable aux algorithmes existants ⁷

⁷A. Ali, M. Meilă, *Experiments with Kemeny ranking: What works when?*, Mathematical Social Science, 64, pp. 28–40, 2012.



Travail publié: résumé



"Heuristic, Branch-and-Bound Solver and Improved Space Reduction for the Median of Permutations Problem" IWOC2017

Résultats préliminaires et travaux futurs





1) Programmation en nombres entiers: Modèle

$$\text{minimiser : } \sum_{i < j, i, j \in \{1, \dots, n\}} R_{ij}(A) x_{ij} \quad (1)$$

sujet aux contraintes:

$$x_{ij} + x_{ji} = 1, \quad i \neq j, i, j \in \{1, \dots, n\} \quad (2)$$

$$x_{ij} + x_{jk} + x_{ki} \geq 1, \quad i \neq j \neq k, i, j, k \in \{1, \dots, n\} \quad (3)$$

$$x_{ij} \in [0, 1] \quad i \neq j, i, j \in \{1, \dots, n\} \quad (4)$$

$$x_{ij} \text{ entier} \quad i \neq j, i, j \in \{1, \dots, n\} \quad (5)$$

où $R_{ij}(A)$ est le nombre de permutations de A où $j \prec i$

où x_{ij} est la variable indiquant $i \prec j$ dans la solution





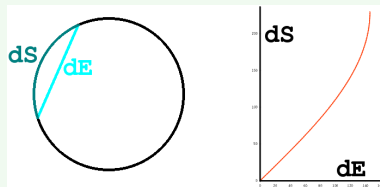
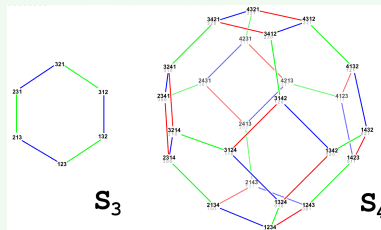
1) Programmation en nombres entiers: CPLEX

taille du problème:		caractéristiques:			
m	n	original	mst	contraints	contraints+mst
3	60	515.56	646.21	449.62	579.77
4	60	454.89	578.87	203.38	332.12
5	60	14288.28	8788.22	9688.74	6183.98
3	70	1563.79	1788.11	1396.94	1606.77
4	70	720.38	909.04	116.75	332.31
5	70	1931.09	2134.26	2063.31	2270.31
3	80	304903.18	143389.36	106089.00	77493.30
4	80	1470.65	1771.20	1209.36	1517.77
5	80	33023.36	4540.35	26113.12	4975.42
3	90	4140.55	4586.23	3475.34	3907.65
4	90	197793.46	15106.97	132881.12	115424.61
5	90	2781696.72	2383961.55	679621.16	908664.74
3	100	775017.20	104621.75	160247.34	37931.90
4	100	4182.20	4807.44	5100.42	5630.65
5	100	136065.61	14990.22	101878.41	15293.40
somme des temps:		4257766.92	2692619.78	1230534.01	1182144.70

Figure: Temps de résolution en ticks de CPLEX pour des instances du problème de $m = 3, 4, 5$ permutations et de taille allant de $n = 60$ à $n = 100$ éléments.



2) Formule d'approximation: Motivation et idée



2) Formule d'approximation: Motivation et idée

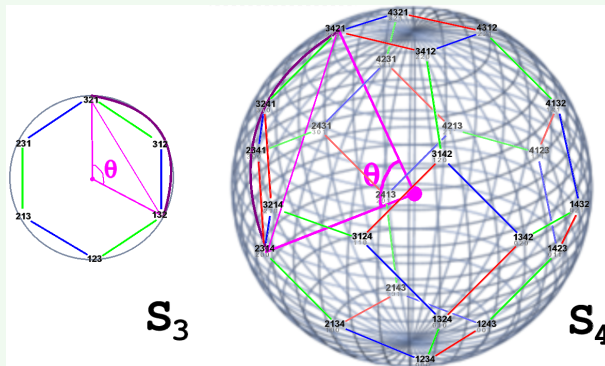


Figure: Permutahedrons pour $n = 3$ et $n = 4$ et sphères superposées.

2) Formule d'approximation: La formule d'approximation

$$d_{K\tau}(\pi, \sigma) \simeq d_S \times \frac{\text{distance maximale de Kendall-}\tau}{\text{distance maximale sur le cercle}}$$

$$d_{K\tau}(\pi, \sigma) \simeq \theta R \times \frac{n(n-1)}{2R\pi}$$

$$\theta = \arccos \left(1 - \frac{d_{eucl}^2}{2R^2} \right) \quad R = \frac{\sqrt{2 \binom{n+1}{3}}}{2}$$

$$d_{K\tau}(\pi, \sigma) \simeq \arccos \left(1 - \frac{\sum_{i \in \{1, \dots, n\}} (\pi_i^{-1} - \sigma_i^{-1})^2}{\binom{n+1}{3}} \right) \times \frac{n(n-1)}{2\pi}$$

2) Formule d'approximation: La formule d'approximation

$$d_{K\tau}(\pi, \sigma) \simeq d_S \times \frac{\text{distance maximale de Kendall-}\tau}{\text{distance maximale sur le cercle}}$$

$$d_{K\tau}(\pi, \sigma) \simeq \theta R \times \frac{n(n-1)}{2R\pi}$$

$$\theta = \arccos \left(1 - \frac{d_{eucl}^2}{2R^2} \right) \quad R = \frac{\sqrt{2 \binom{n+1}{3}}}{2}$$

$$d_{K\tau}(\pi, \sigma) \simeq \arccos \left(1 - \frac{\sum_{i \in \{1, \dots, n\}} (\pi_i^{-1} - \sigma_i^{-1})^2}{\binom{n+1}{3}} \right) \times \frac{n(n-1)}{2\pi}$$

2) Formule d'approximation: La formule d'approximation

$$d_{K\tau}(\pi, \sigma) \simeq d_S \times \frac{\text{distance maximale de Kendall-}\tau}{\text{distance maximale sur le cercle}}$$

$$d_{K\tau}(\pi, \sigma) \simeq \theta R \times \frac{n(n-1)}{2R\pi}$$

$$\theta = \arccos \left(1 - \frac{d_{eucl}^2}{2R^2} \right) \quad R = \frac{\sqrt{2 \binom{n+1}{3}}}{2}$$

$$d_{K\tau}(\pi, \sigma) \simeq \arccos \left(1 - \frac{\sum_{i \in \{1, \dots, n\}} (\pi_i^{-1} - \sigma_i^{-1})^2}{\binom{n+1}{3}} \right) \times \frac{n(n-1)}{2\pi}$$

2) Formule d'approximation: La formule d'approximation

$$d_{K\tau}(\pi, \sigma) \simeq d_S \times \frac{\text{distance maximale de Kendall-}\tau}{\text{distance maximale sur le cercle}}$$

$$d_{K\tau}(\pi, \sigma) \simeq \theta R \times \frac{n(n-1)}{2R\pi}$$

$$\theta = \arccos \left(1 - \frac{d_{eucl}^2}{2R^2} \right) \quad R = \frac{\sqrt{2 \binom{n+1}{3}}}{2}$$

$$d_{K\tau}(\pi, \sigma) \simeq \arccos \left(1 - \frac{\sum_{i \in \{1, \dots, n\}} (\pi_i^{-1} - \sigma_i^{-1})^2}{\binom{n+1}{3}} \right) \times \frac{n(n-1)}{2\pi}$$

2) Formule d'approximation: Résultats

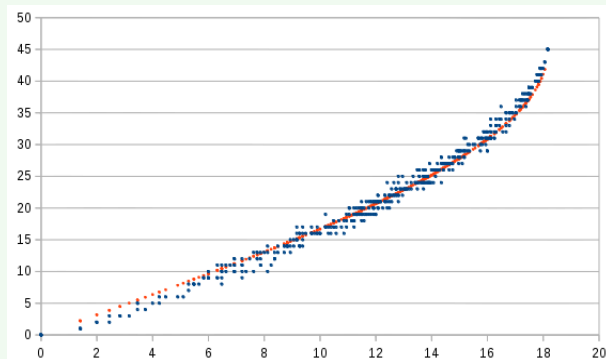
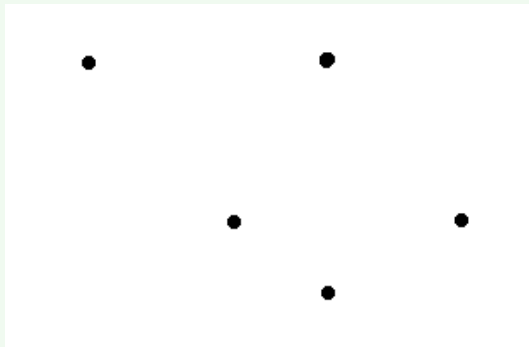
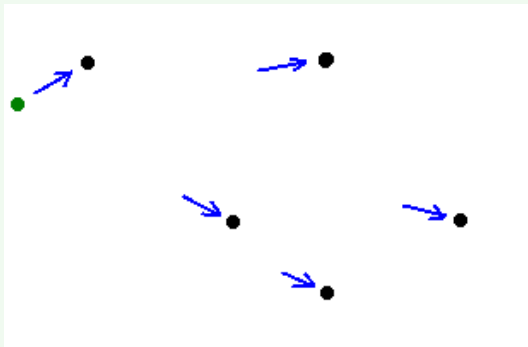


Figure: Distance de Kendall- τ en fonction de la distance euclidienne pour des permutations de $n = 10$. Le taux d'écart de 6.00% en moyenne, stat. sur 500 paires de permutations aléatoires.

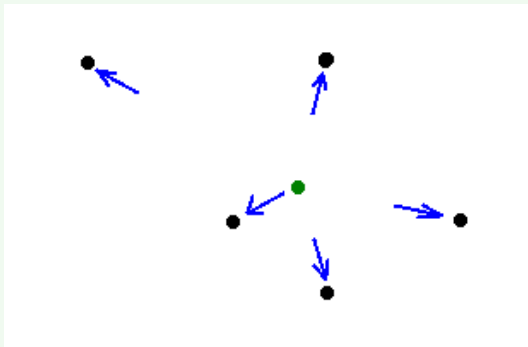
3) Heuristique "MedianGame": Motivation



3) Heuristique "MedianGame": Motivation



3) Heuristique "MedianGame": Motivation



3) Heuristique "MedianGame": Exemple

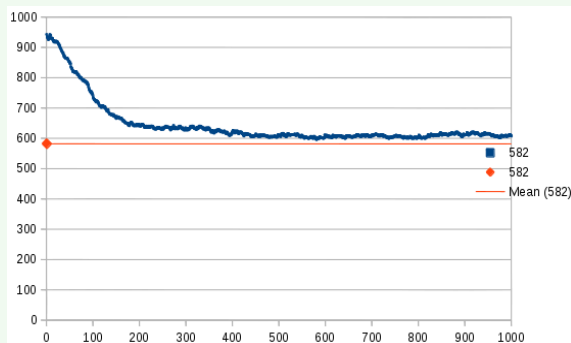


Figure: Score de la solution courante à travers les itérations de l'heuristique Median Game pour une instance de $m = 4$ permutations de taille $n = 30$. Le minimum enregistré par l'heuristique est de 596 (soit à 2.41% de l'optimal 582).

4) Grandes instances: Idées

Grandes instances

- heuristiques: recuit simulé, MedianGame et descente de gradient de la formule d'approximation
- optimisation locale: "window-optimizer" (k -opt)
- bornes inférieures: Conitzer⁸ et programmation linéaire

⁸V. Conitzer, A. Davenport, and J. Kalagnanam, *Improved bounds for computing Kemeny rankings*, Proceedings of the 21st conference on Artificial intelligence - Volume 1, AAAI'06, pp. 620–626, 2006.



5) Classements avec égalité: Contexte

BI = [[TPH2],[DRD5,DRD4],[HTR1B,SLC6A2,SLC6A3,SNAP25,ADRA2A],[ADHD],[FGD1,SLITRK1],[SLC9A9],...
 IE = [[TPH2],[HTR1B,SLC6A2,SLC6A3,SNAP25,ADRA2A,DRD5,DRD4],[ADHD],[TPH1,CES1],[RAI1,SLC9A9],...
 PR = [[TPH2],[HTR1B,SLC6A2,SLC6A3,SNAP25,ADRA2A,DRD4],[DRD5],[FGD1],[ADHD],[SLC9A9,TPH1],...
 PC = [[TPH2],[DRD5,DRD4],[HTR1B,SLC6A2,SLC6A3,SNAP25,ADRA2A],[ADHD],[TPH1],[FGD1,SLITRK1],...

mapping



consensus



$R_1 = [[7],[3,2],[31,41,4,5,1],[8],[27,43],[42],...$
 $R_2 = [[7],[31,41,4,5,1,3,2],[8],[6,17],[38,42],...$
 $R_3 = [[7],[31,41,4,5,1,2],[3],[27],[8],[42,6],...$
 $R_4 = [[7],[3,2],[31,41,4,5,1],[8],[6],[27,43],...$



$C = [[7],[2,3],[1,4,5,31,41],[8],[6],[27,43],...$

mapping



$C = [[TPH2],[DRD4,DRD5],[ADRA2A,SLC6A3,SNAP25,HTR1B,SLC6A2],[ADHD],[TPH1],[FGD1,SLITRK1],...$

Figure: Exemple de classements de gènes pour la maladie de l'ADHD. Données provenant du site de BioConcert
<http://bioguide-project.net/bioconsert/>.



5) Classements avec égalité: Transposer les connaissances

- les heuristiques
- les contraintes
- les bornes inférieures Conitzer et extensions
- la programmation en nombre entiers et l'utilisation de CPLEX
- le calcul rapide de la distance de Kendall- τ généralisée

Calendrier

session	objectifs
A-2017	programmation en nombres entiers, formule d'approximation, article journal (Stage Mitacs) classements avec égalité, grandes instances, bio-info
H-2018	(Stage Mitacs) rédaction article, rapport Mitacs, classements avec égalité, début organisation thèse
E-2018	programmation en nombres entiers, rédaction thèse, rédaction article
A-2018	dépôt de la thèse, correction et soutenance



Merci

Questions?