Signal Processing on the Permutahedron

Tight Spectral Frames for Ranked Data Analysis

David Shuman

Joint work with Yilin (Ellen) Chen, Jennifer DeJong, and Tom Halverson

November 7, 2022

LearnGraph '22

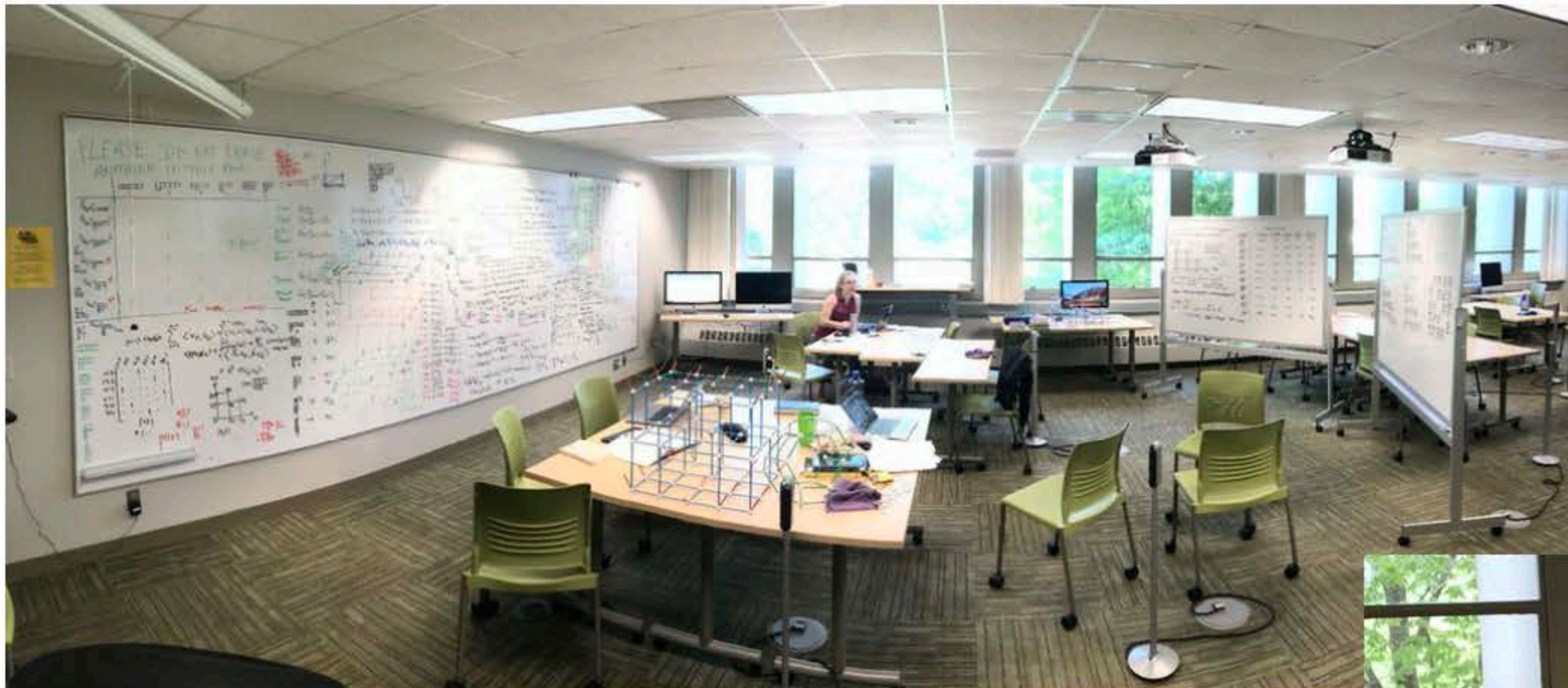
CIRM, Marseille, France

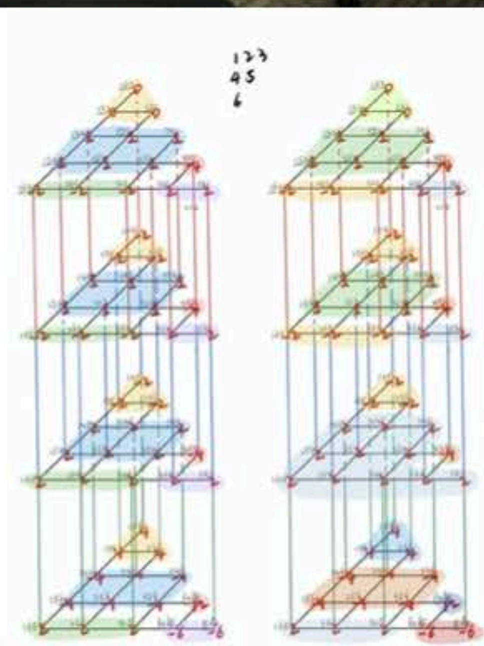
These slides are available at bit.ly/sp-perm

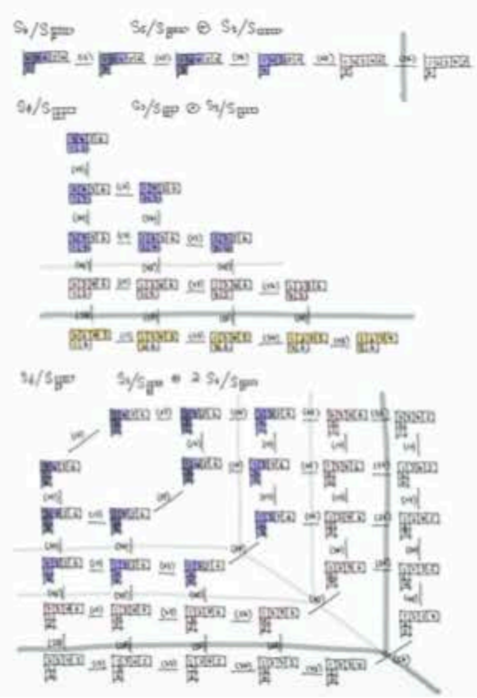
Olin College of Engineering

MACALESTER

Undergraduate Research Collaborators





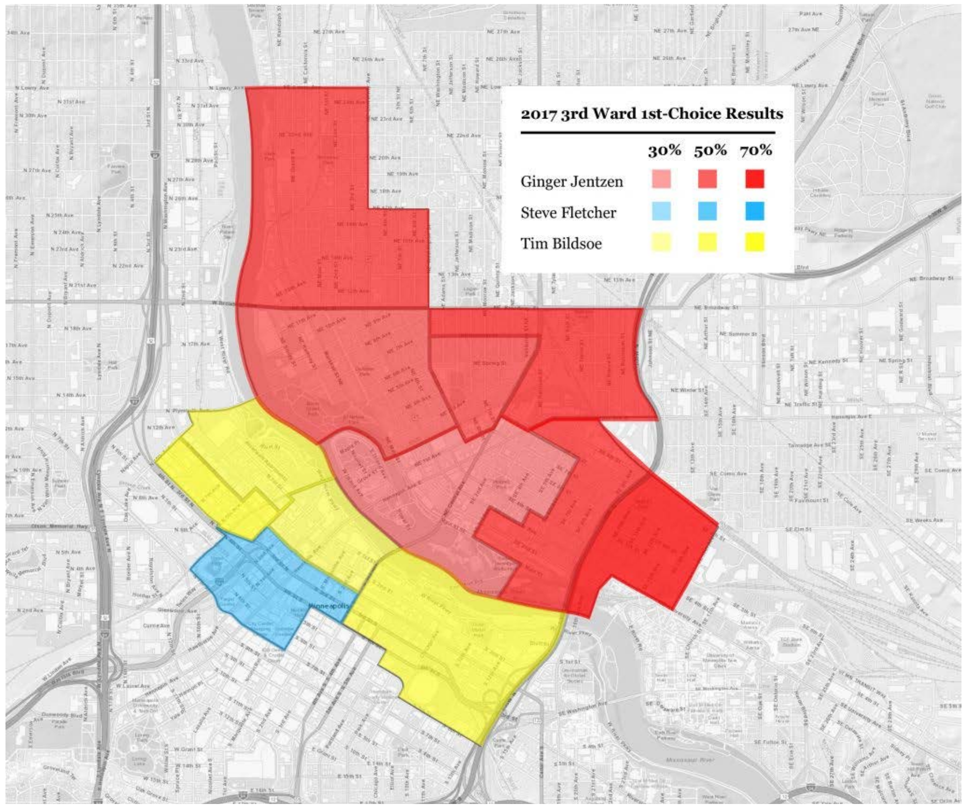


[Chen, DeJong, Halverson, and Shuman, "Signal Processing on the Permutahedron: Tight Spectral Frames for Ranked Data Analysis," Journal of Fourier Analysis and Applications, Aug. 2021]

Ranked Data Example:
2017 Minneapolis City Council Ward 3 Election

Four Candidates:

1. Ginger Jentzen
   (Socialist-Alternative)
2. Samantha Pree-Stinson
   (Green)
3. Steven Fletcher
   (Democratic-Farmer-Labor, elected)
4. Tim Bildsoe
   (Democratic-Farmer-Labor)



1st	2nd	3rd	4th	Count
4	3	2	1	574
4	3	1	2	201
4	2	3	1	131
4	2	1	3	32
4	1	3	2	89
4	1	2	3	46
3	4	2	1	422
3	4	1	2	271
3	2	4	1	159
3	2	1	4	243
3	1	4	2	156
3	1	2	4	204
2	4	3	1	111
2	4	1	3	30
2	3	4	1	161
2	3	1	4	145
2	1	4	3	56
2	1	3	4	153
1	4	3	2	255
1	4	2	3	77
1	3	4	2	376
1	3	2	4	421
1	2	4	3	204
1	2	3	4	538

SP>GJ>SF>TB

https://streets.mn/2019/05/20/how-the-2017-ward-3-election-in-minneapolis-foreshadows-our-local-political-future/

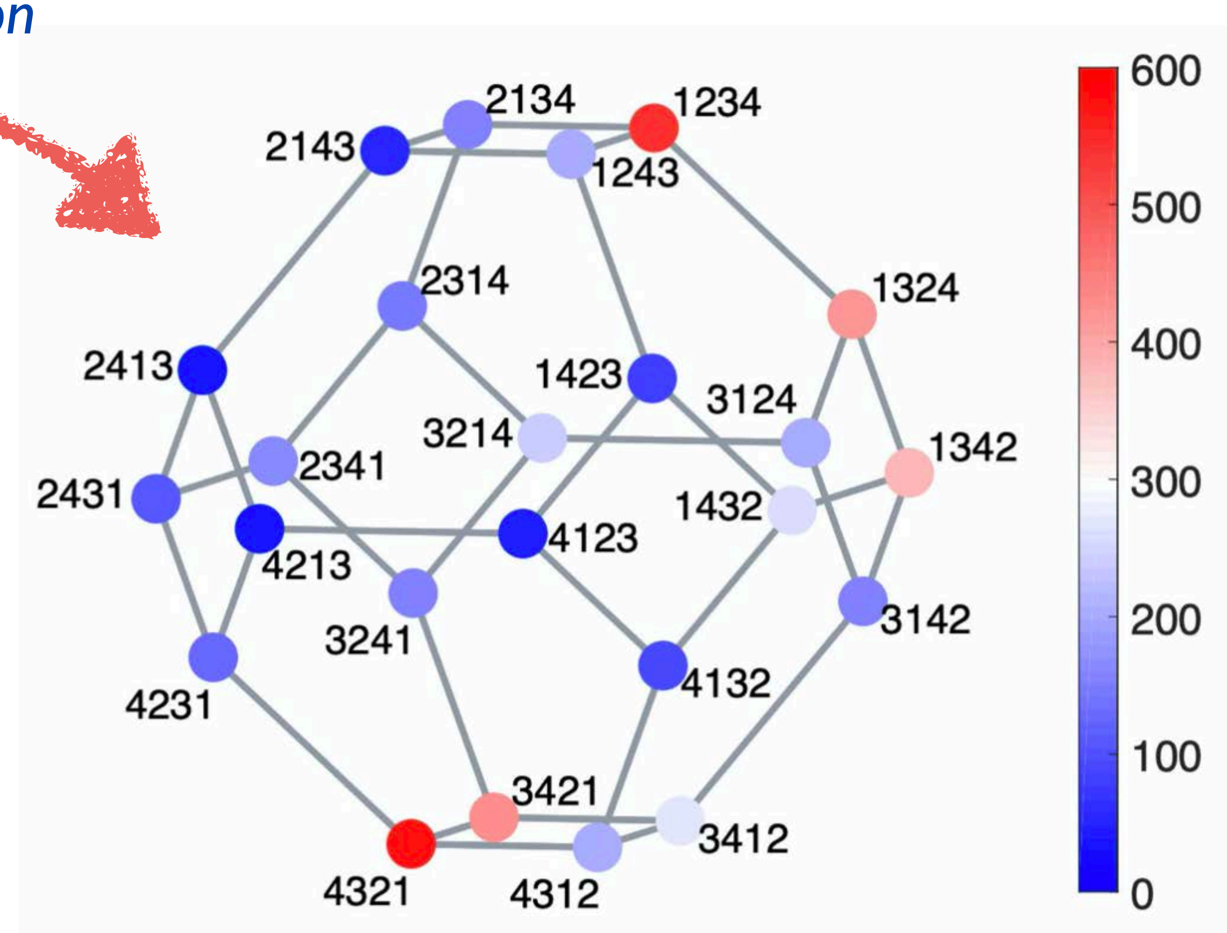
Ranked Data in Other Applications

- Crowdsourcing knowledge
  - Bioinformatics/genomics: combining results of microarray experiments (e.g., to identify genes of interest)
  - Peer grading
  - Crowdsourced subjective labeling
- Computer vision and image processing
  - Multi-object tracking / air traffic control
  - Photosequencing
- Other analytics
  - Lineup composition and utilization for sports analytics (which players work well together?)
  - Similarities between Supreme Court justices (each case is a voter and it selects, e.g., the judges in the minority opinion)

Ranked Data Lives on the Permutahedron

2017 Minneapolis City Council Ward 3 Election

1st	2nd	3rd	4th	Count
4	3	2	1	574
4	3	1	2	201
4	2	3	1	131
4	2	1	3	32
4	1	3	2	89
4	1	2	3	46
3	4	2	1	422
3	4	1	2	271
3	2	4	1	159
3	2	1	4	243
3	1	4	2	156
3	1	2	4	204
2	4	3	1	111
2	4	1	3	30
2	3	4	1	161
2	3	1	4	145
2	1	4	3	56
2	1	3	4	153
1	4	3	2	255
1	4	2	3	77
1	3	4	2	376
1	3	2	4	421
1	2	4	3	204
1	2	3	4	538



$$f : \mathbb{S}_n \to \mathbb{R}.$$

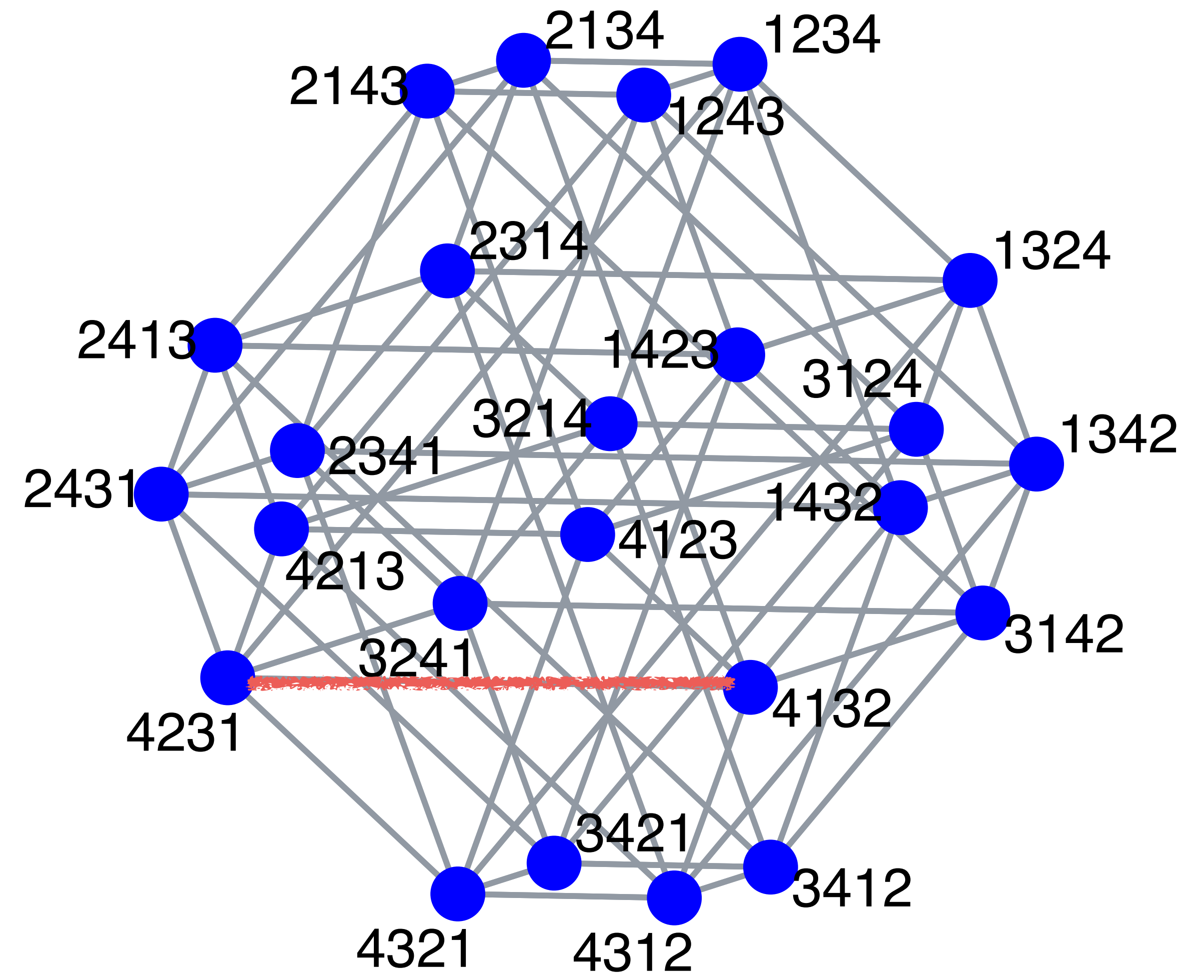
- The permutahedron is the Cayley graph of the symmetric group generated by adjacent transpositions
- Rankings that differ by a single swap of neighboring candidates are close from a voter's viewpoint

Why is the Permutahedron the Appropriate Graph for Most Ranked Data Analyses?

- Alternative choice: Cayley graph of the symmetric group induced by the generating set of *all* transpositions (not just neighboring transpositions)

- This Cayley graph captures the appropriate notion of distance when the "slots" do not represent linearly ordered rankings
  - In multi-object tracking, associate an object with a trajectory

- This graph has many nice theoretical properties, e.g., eigenvalues and eigenvectors are known in closed form



Kondor, Howard, Jebara, "Multi-object tracking with representations of the symmetric group," AISTAT, 2007.

Ghandehari, Guillot, Hollingsworth, "A non-commutative viewpoint on graph signal processing," SAMPTA, 2019

Research Questions & Approach

Main research questions: How do we identify, interpret, and exploit structure in ranked data?

Two main classes of approaches in the literature:

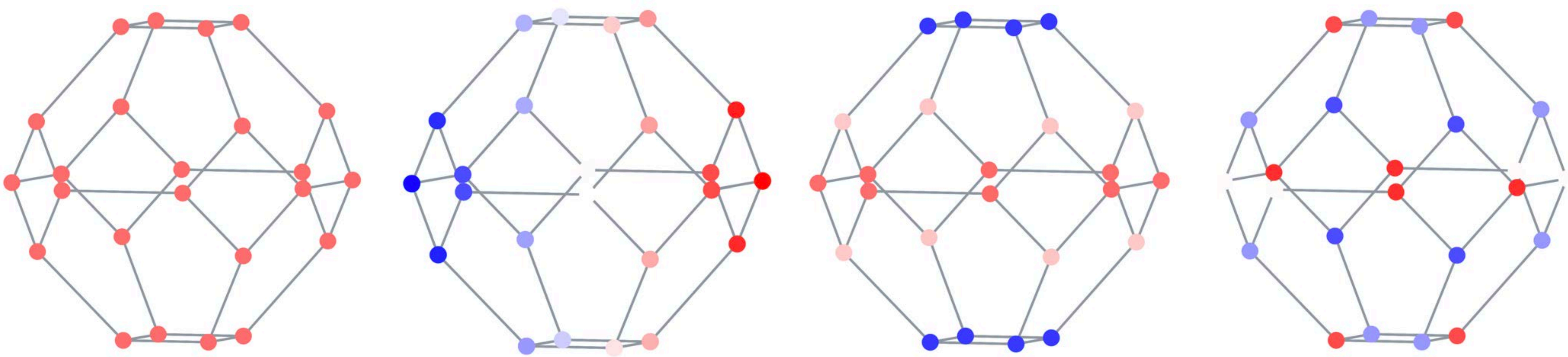
1. Parametric statistical models
   - Order statistic models
   - Distance-based models
   - Pairwise comparison models
2. Linear transforms
   - Nested orthogonal contrasts
   - Inversions
   - Fourier analysis on the symmetric group

Marden, Analyzing and Modeling Rank Data, 1995

Yu, Gu, and Xu, "Analysis of ranking data," Wiley Interdisciplinary Reviews, 2019

The Graph Signal Processing Approach
Smoothness with Respect to the Graph

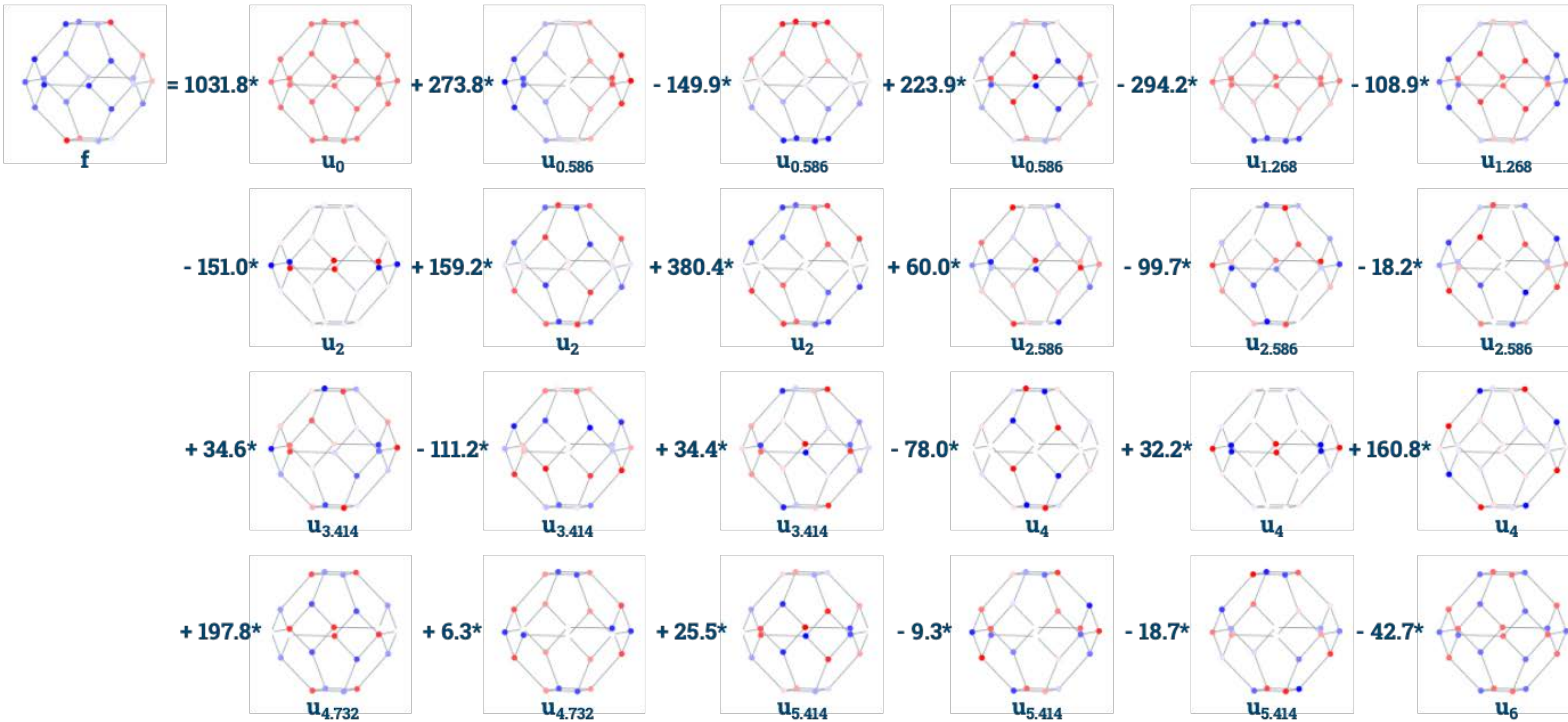
- Decompose a signal into a linear combination of orthonormal graph Laplacian eigenvectors
- $\mathbb{R}[\mathbb{S}_n] \cong \bigoplus_{\lambda} U_\lambda$
- Eigenvectors capture a notion of smoothness with respect to the graph
- Values of the Laplacian eigenvectors associated with low eigenvalues change less rapidly across connected vertices



$\lambda = 0$ $\lambda = 0.586$ $\lambda = 1.268$ $\lambda = 4.732$

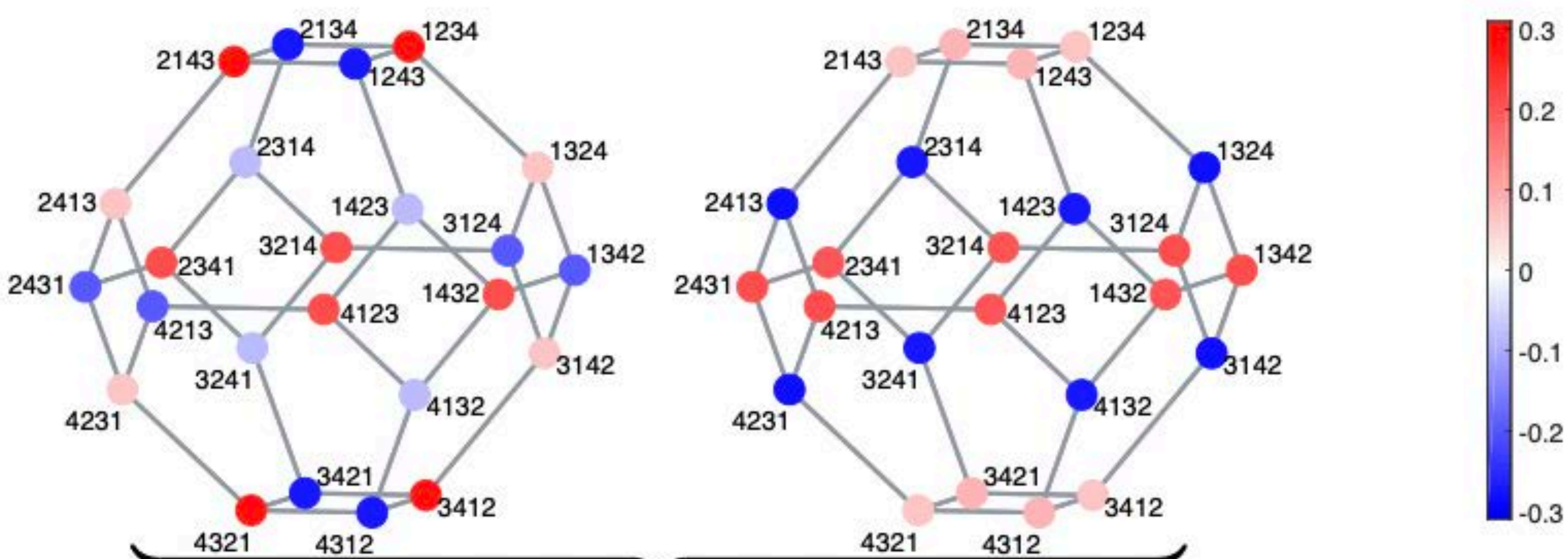
The Graph Signal Processing Approach
Example GFT

2017 Minneapolis City Council Ward 3 Election



The Graph Signal Processing Approach Shortcomings

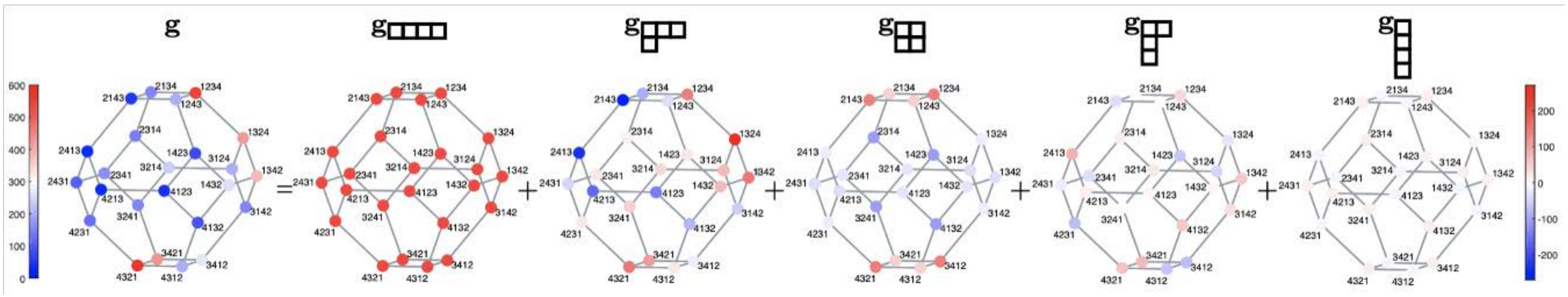
- Due to the symmetries of the permutahedron, there are many repeated Laplacian eigenvalues
- We can get the energy of a signal on a given eigenspace with repeated eigenvalues, but it is not clear how to specify a basis for each eigenspace that captures interesting symmetries
- More broadly, the general GSP approach does not explicitly take advantage of the symmetry of the domain



Orthornomal basis for $U_{4.7321}$

Group Representation Theory Approach
Symmetry Decomposition & Isotypic Components

- $\mathbb{R}[\mathbb{S}_n] \cong \bigoplus_{\gamma \vdash n} W_\gamma$: Decomposition into orthogonal subspaces called isotypic components
- Indexed by integer partitions of n (also called *shape* or *symmetry type*)
- Isotypic components can be further decomposed into irreducible submodules



$$1{,}607{,}249.0 = \underbrace{1{,}064{,}709.4 + 355{,}201.6 + 137{,}575.3 + 47{,}942.6 + 1820.0}_{\textbf{Energy Decomposition}}$$

Group Representation Theory Approach
Relation to Marginal Statistics

First Order Marginals

Candidate	First Choice	Second Choice	Third Choice	Fourth Choice
Ginger Jentzen	1871	704	922	1558
Samantha Pree-Stinson	656	1307	1744	1348
Steve Fletcher	1455	1878	1277	445
Tim Bildsoe	1073	1166	1112	1704

- The isotypic component $W_{(4)}$ captures zeroth order information about the total number of voters
- The isotypic component $W_{(3,1)}$ captures first order marginals, net of zeroth order marginals

Second Order Unordered Marginals

Candidates	First and Second	First and Third	First and Fourth	Second and Third	Second and Fourth	Third and Fourth
Ginger Jentzen and Samantha Pree-Stinson	951	673	903	525	535	1468
Ginger Jentzen and Steve Fletcher	1157	1307	862	588	837	304
Ginger Jentzen and Tim Bildsoe	467	813	1664	513	890	708
Samantha Pree-Stinson and Steve Fletcher	708	890	513	1664	813	467
Samantha Pree-Stinson and Tim Bildsoe	304	837	588	862	1307	1157
Steve Fletcher and Tim Bildsoe	1468	535	525	903	673	951

- $W_{(2,2)}$ captures second order unordered marginals, net of zeroth and first order marginals

Second Order Ordered Marginals

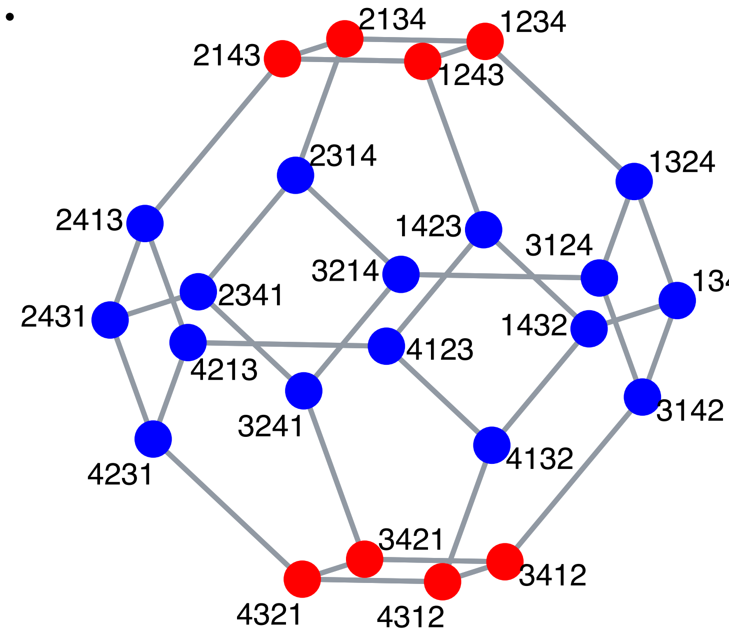
Candidate1	Candidate2	First-Second	First-Third	First-Fourth	Second-Third	Second-Fourth	Third-Fourth
Ginger Jentzen	Samantha Pree-Stinson	742	498	631	250	245	472
Ginger Jentzen	Steve Fletcher	797	793	281	242	102	62
Ginger Jentzen	Tim Bildsoe	332	580	959	212	357	388
Samantha Pree-Stinson	Ginger Jentzen	209	175	272	275	290	996
Samantha Pree-Stinson	Steve Fletcher	306	264	86	669	236	123
Samantha Pree-Stinson	Tim Bildsoe	141	217	298	363	781	625
Steve Fletcher	Ginger Jentzen	360	514	581	346	735	242
Steve Fletcher	Samantha Pree-Stinson	402	626	427	995	577	344
Steve Fletcher	Tim Bildsoe	693	315	447	537	566	691
Tim Bildsoe	Ginger Jentzen	135	233	705	301	533	320
Tim Bildsoe	Samantha Pree-Stinson	163	620	290	499	526	532
Tim Bildsoe	Steve Fletcher	775	220	78	366	107	260

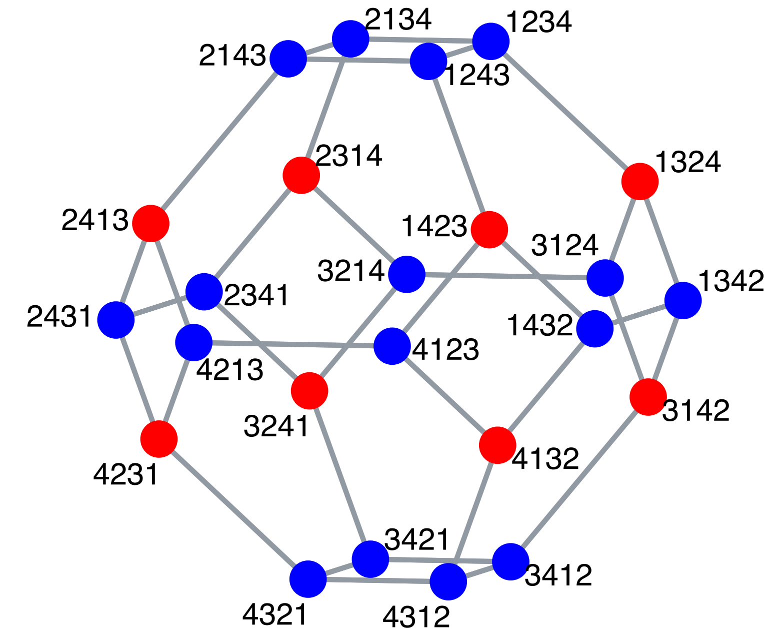
- $W_{(2,1,1)}$ captures second order ordered marginals, net of zeroth and first order marginals, and net of second order unordered marginals

Summary: Each isotypic component captures the corresponding marginals, net of all the isotypic components preceding it in *dominance order*

Group Representation Theory Approach Shortcomings

- Two (quite different) signals with the exact same energy decomposition into isotopic components:





γ	(4)	$(3,1)$	$(2,2)$	$(2,1,1)$	$(1,1,1,1)$
$\Vert\mathbf{f}_\gamma\Vert^2$	2.67	0	5.33	0	0

- More refined approach: Fourier transform on the symmetric group
  - Well known issue: no natural choice of basis for each subspace
  - Common workaround: choose a basis that is well-adapted to computation, but lacks interpretability

Diaconis, "A generalization of spectral analysis with application to ranked data," The Annals of Statistics, 1989

Outline

- Background on ranked choice voting, signals on the permutahedron, and group representation theory
- **Design of tight spectral frames for ranked data analysis**
- Illustrative analysis examples
- Computational challenges and efficient algorithms
- Extensions and ongoing work

Our Approach: Combine the Spectral and Symmetry Decompositions

First step:

Show that the signal space always decomposes into subspaces that are the intersections of the Laplacian eigenspaces and the isotopic components:

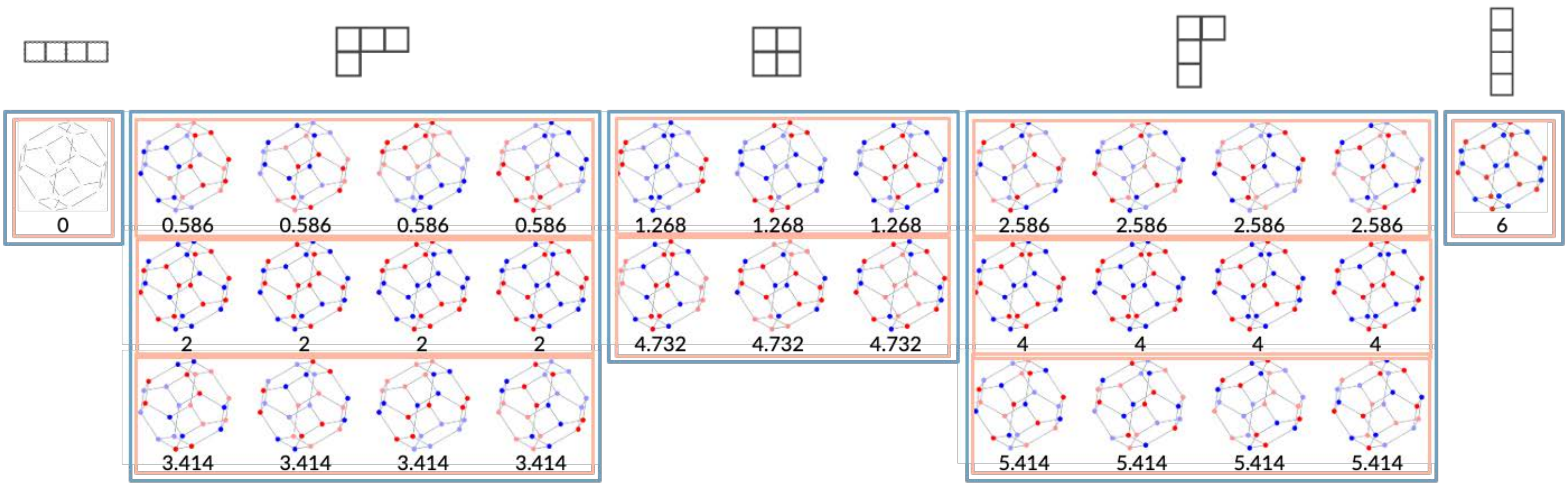
$$\mathbb{R}[\mathbb{S}_n] \cong \bigoplus_{\gamma \vdash n} \bigoplus_{\lambda \in \Lambda_\gamma} Z_{\gamma,\lambda}, \quad \text{where} \quad Z_{\gamma,\lambda} = W_\gamma \cap U_\lambda$$

Dictionary design objective:

For each space $Z_{\gamma,\lambda}$, find a spanning set of dictionary atoms (vectors) with interpretable patterns that capture both **smoothness** and **structural** information about the ranked data on the permutahedron

Example of a Dictionary for $\mathbb{R}[\mathbb{S}_4]$

$$\mathbb{R}[\mathbb{S}_n] \cong \bigoplus_{\gamma \vdash n} \bigoplus_{\lambda \in \Lambda_\gamma} Z_{\gamma,\lambda}, \quad \text{where} \quad Z_{\gamma,\lambda} = W_\gamma \cap U_\lambda$$

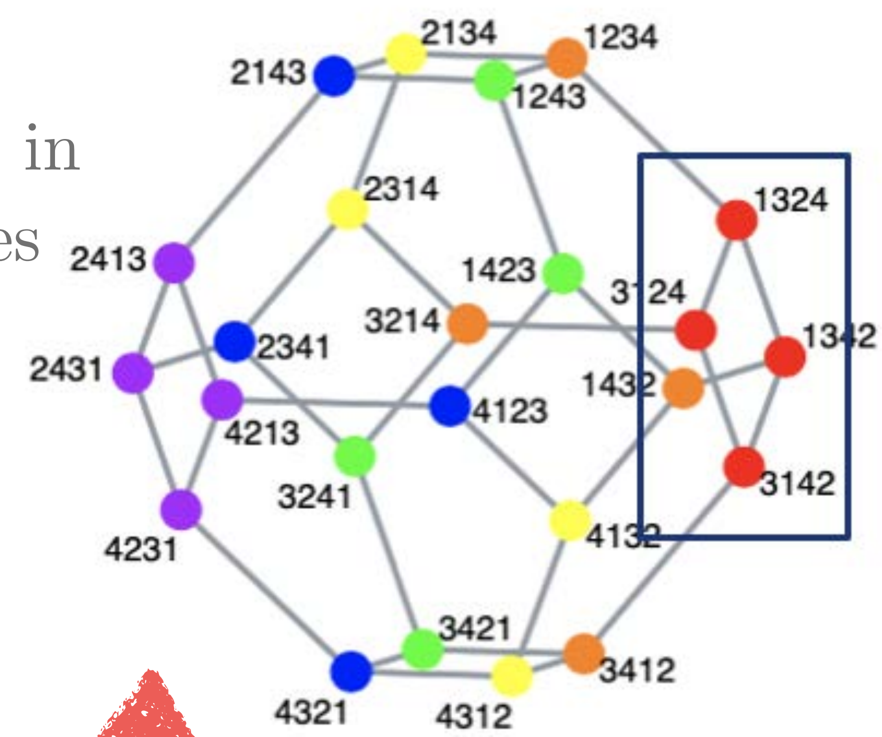


- This particular dictionary (4 candidates) has 32 atoms that span a 24-dimensional vector space
- In this example, the Laplacian eigenvectors associated with any particular eigenvalue belong to a single isotopic component, but that is not the case in general (e.g., $\lambda = 3$ appears in multiple shapes with 6 candidates)

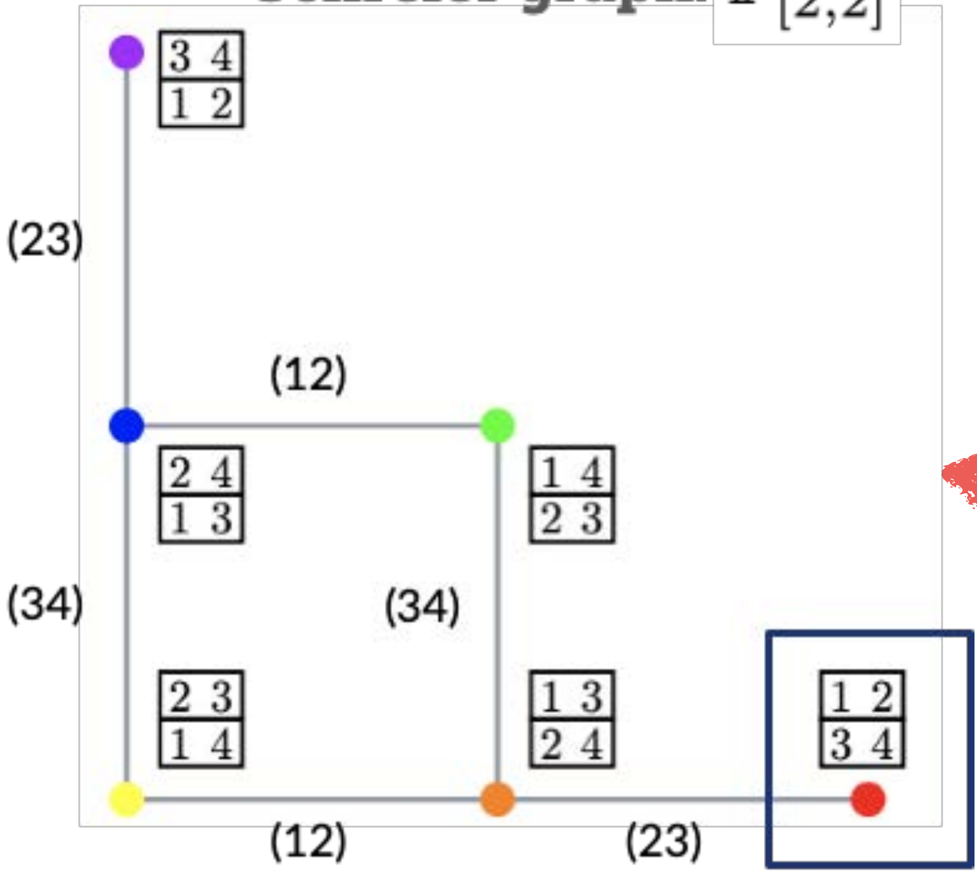
Background: Equitable Partitions & Schreier Graphs

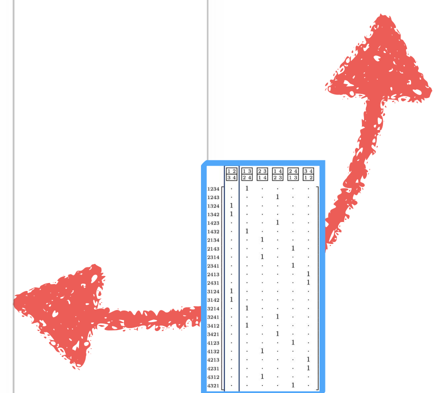
Example equitable partition:

Group vertices with candidates 1 & 3 in the same ranking slots, and candidates 2 & 4 in the same ranking slots



Schreier graph: $\mathbb{P}_{[2,2]}$





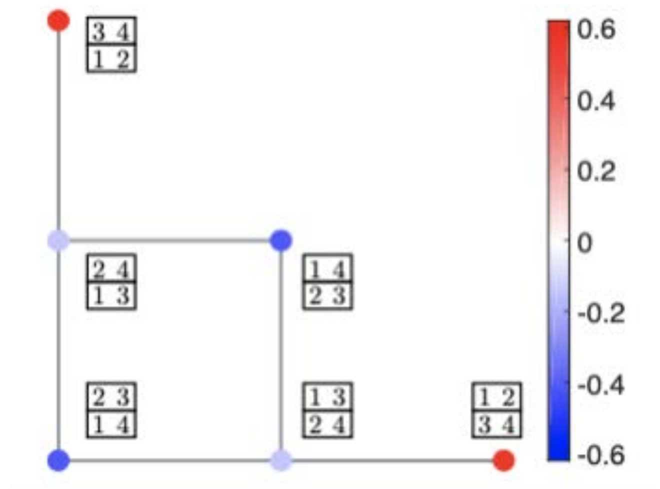
	1 2 / 3 4	1 3 / 2 4	2 3 / 1 4	1 4 / 2 3	2 4 / 1 3	3 4 / 1 2
1234	·	1	·	·	·	·
1243	·	·	·	1	·	·
1324	1	·	·	·	·	·
1342	1	·	·	·	·	·
1423	·	·	·	1	·	·
1432	·	1	·	·	·	·
2134	·	·	1	·	·	·
2143	·	·	·	·	1	·
2314	·	·	1	·	·	·
2341	·	·	·	·	1	·
2413	·	·	·	·	·	1
2431	·	·	·	·	·	1
3124	1	·	·	·	·	·
3142	1	·	·	·	·	·
3214	·	1	·	·	·	·
3241	·	·	·	1	·	·
3412	·	1	·	·	·	·
3421	·	·	·	1	·	·
4123	·	·	·	·	1	·
4132	·	·	1	·	·	·
4213	·	·	·	·	·	1
4231	·	·	·	·	·	1
4312	·	·	1	·	·	·
4321	·	·	·	·	1	·

Vertices = ordered set partitions of a fixed shape (integer partition)
[numbers correspond to ranking slots, not candidates]

Edges = two adjacent slot numbers are swapped

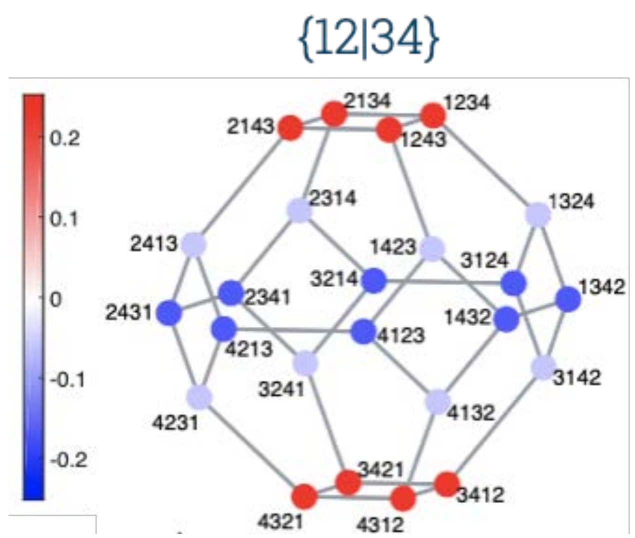
Tight Frame for One Symmetry-Eigen Subspace $Z_{\gamma,\lambda} = W_{[2,2]} \cap U_{1.2679}$

1. Compute a Laplacian eigenvector of the Schreier graph

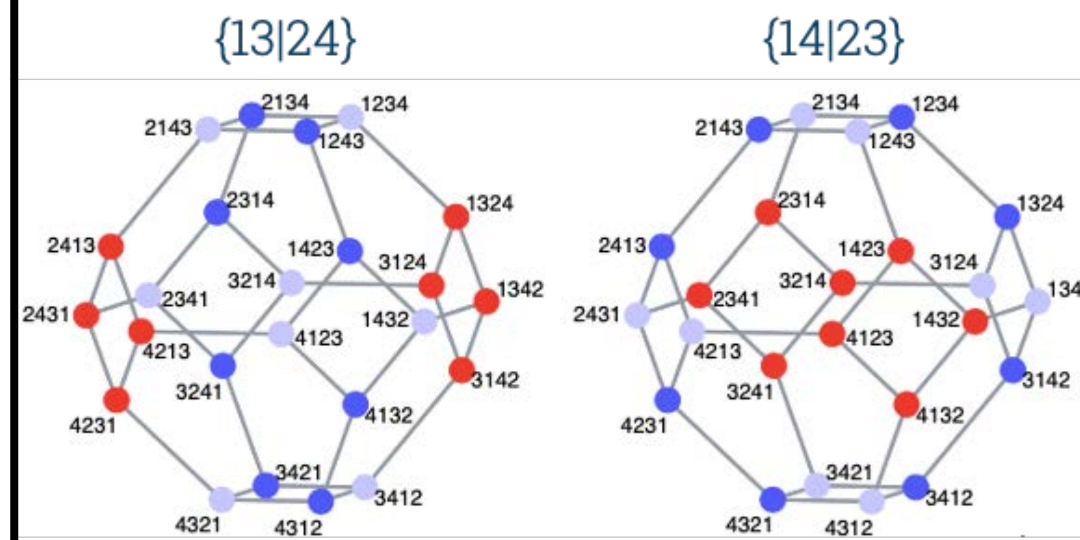


$\gamma = [2,2], \lambda = 1.2679$

2. Lift to the permutahedron by assignment of candidates



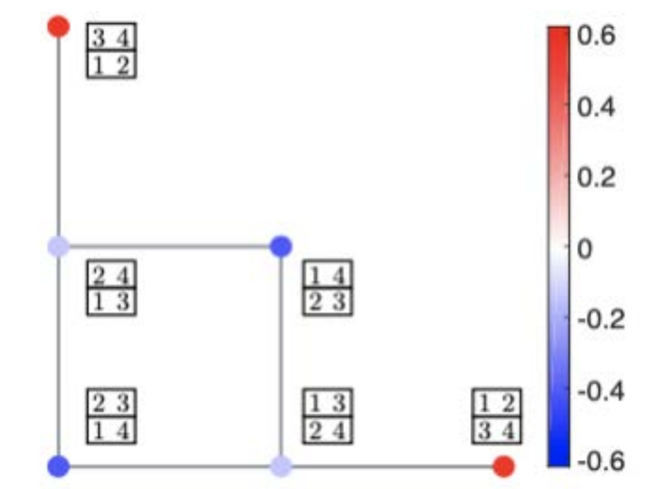
3. Rotate by group elements to obtain more dictionary atoms



- Can also interpret each rotated frame vector as lifting according to a different grouping of candidates

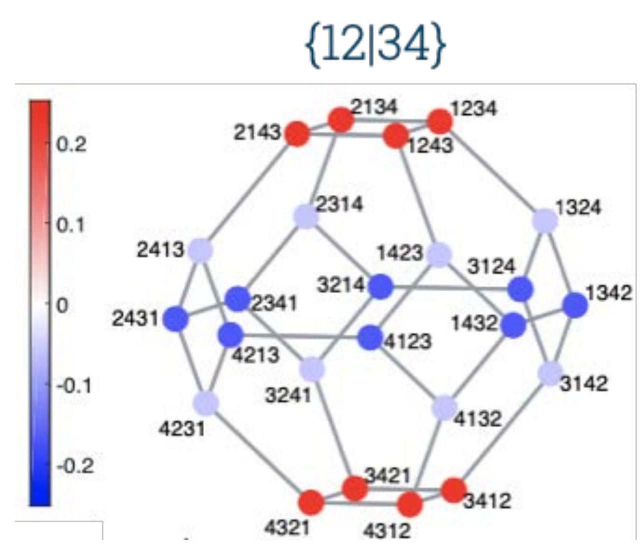
Tight Frame for One Symmetry-Eigen Subspace $Z_{\gamma,\lambda} = W_{[2,2]} \cap U_{1.2679}$

1. Compute a Laplacian eigenvector of the Schreier graph

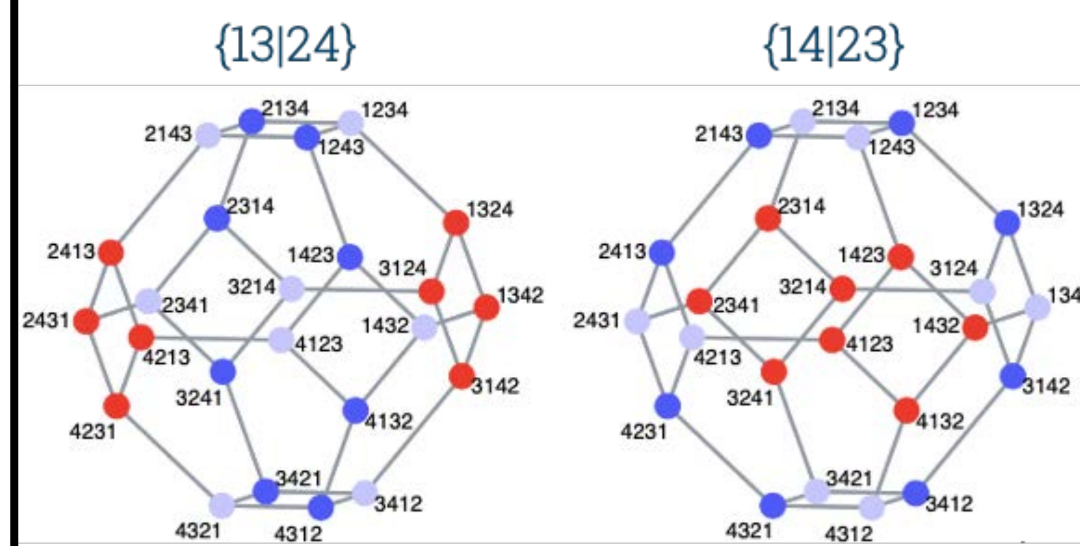


$\gamma = [2,2], \lambda = 1.2679$

2. Lift to the permutahedron by assignment of candidates



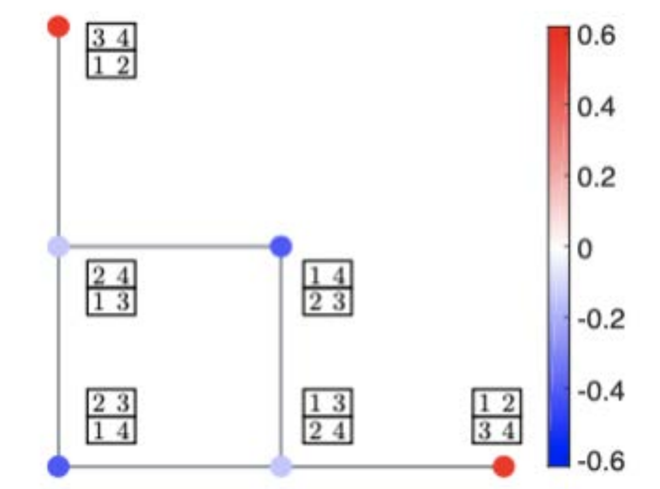
3. Rotate by group elements to obtain more dictionary atoms



- **Why are we constructing atoms in this way?**
  - Fun fact: Multiplying a Laplacian eigenvector of the Schreier graph on the left by the characteristic matrix of an equitable partition yields a Laplacian eigenvector of the permutahedron with the same eigenvalue!!
  - The lifted eigenvectors are also guaranteed to be in the same isotypic component
  - These atoms (and thus the resulting analysis coefficients / inner products with the graph signal) are more interpretable than the numerically computed Laplacian eigenvectors

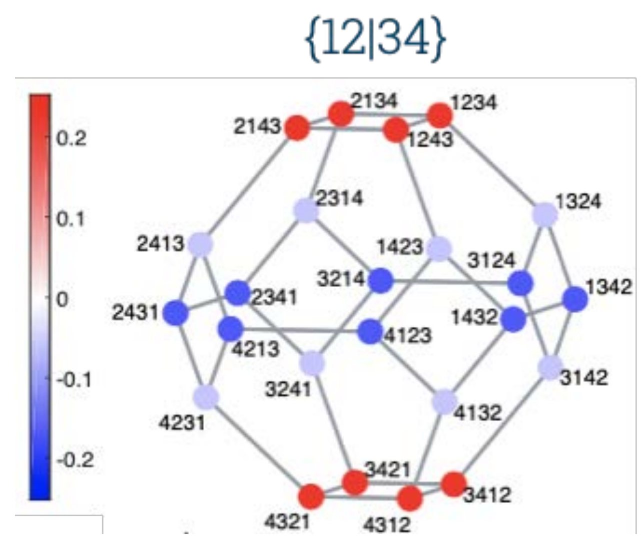
Tight Frame for One Symmetry-Eigen Subspace $Z_{\gamma,\lambda} = W_{[2,2]} \cap U_{1.2679}$

1. Compute a Laplacian eigenvector of the Schreier graph

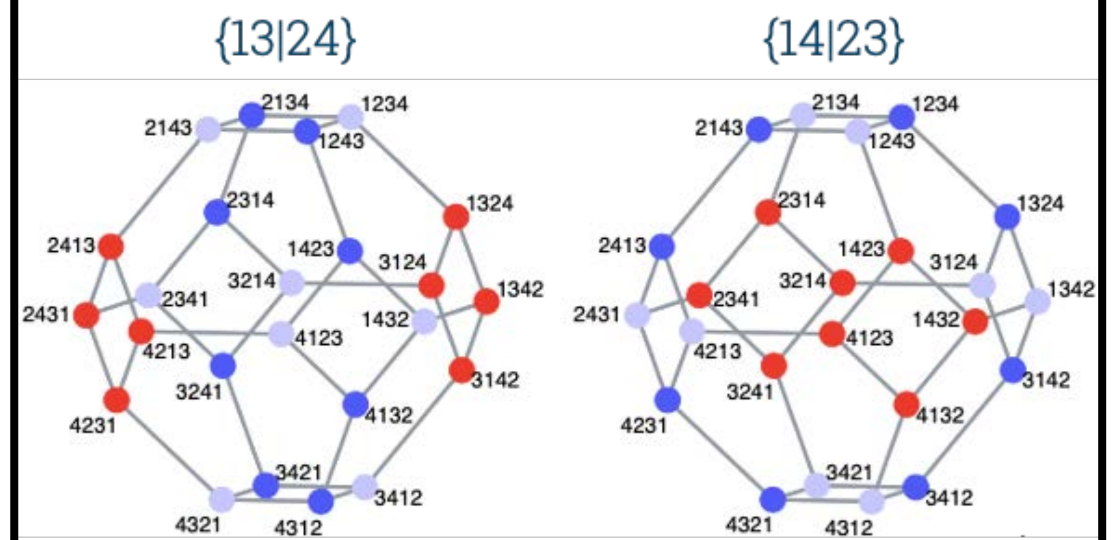


$\gamma = [2, 2], \lambda = 1.2679$

2. Lift to the permutahedron by assignment of candidates



3. Rotate by group elements to obtain more dictionary atoms

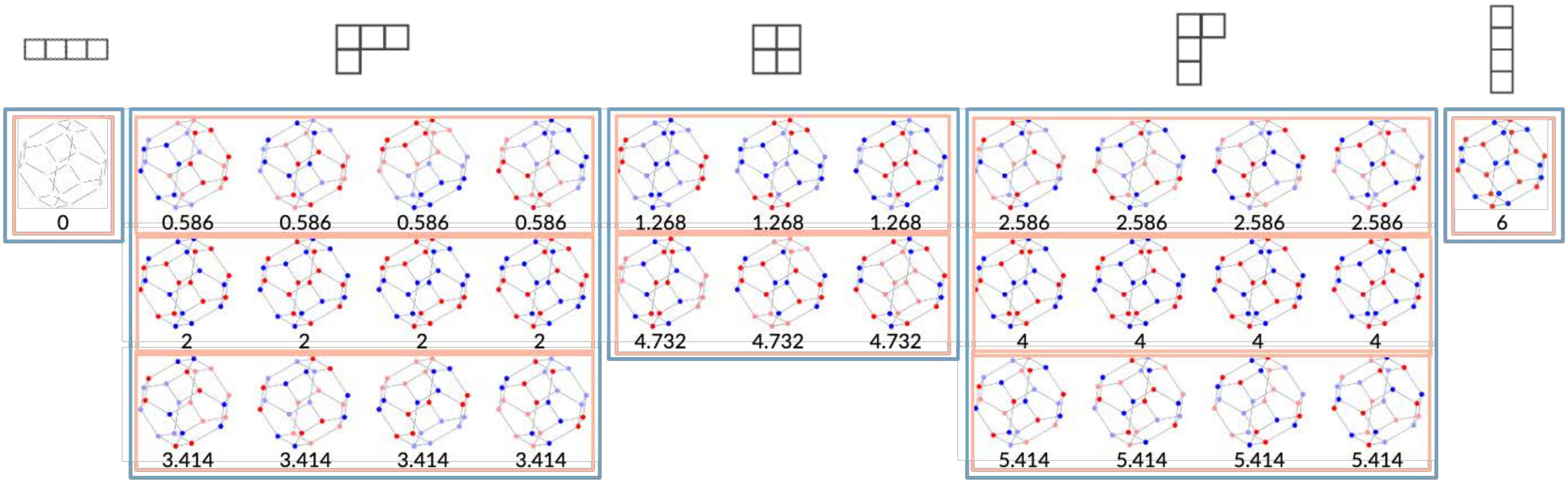


Tight frame for $U_{1.2679}$

- **Best of all:**
  - The dictionary atoms constructed in this manner form a tight frame for $Z_{\gamma,\lambda}$

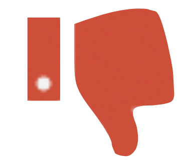
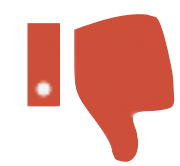
Overall Tight Frame Construction
Example of a Dictionary for $\mathbb{R}[\mathbb{S}_4]$

$$\mathbb{R}[\mathbb{S}_n] \cong \bigoplus_{\gamma \vdash n} \bigoplus_{\lambda \in \Lambda_\gamma} Z_{\gamma,\lambda}, \quad \text{where} \quad Z_{\gamma,\lambda} = W_\gamma \cap U_\lambda$$



- The union of tight Parseval frames for orthogonal subspaces is a tight Parseval frame for their orthogonal sum
- Thus, we can easily build dictionaries that are tight frames for any isotypic component W_γ or the overall signal space

Tight Frame Construction: Rapid Fire Q&A

1. Do you have to lift each eigenvector for each Schreier graph? 
2. Is there a way to know which ones to lift? 
3. Are you going to explain it to us? 
4. Does it involve Kostka numbers, irreducible right submodules, and dominance order? 
5. Is there a slightly crazy formula for how many total atoms you will end up with? 
6. Do you also know in advance the angles between the atoms of each subframe? 
7. Does it involve another crazy formula? 

Outline

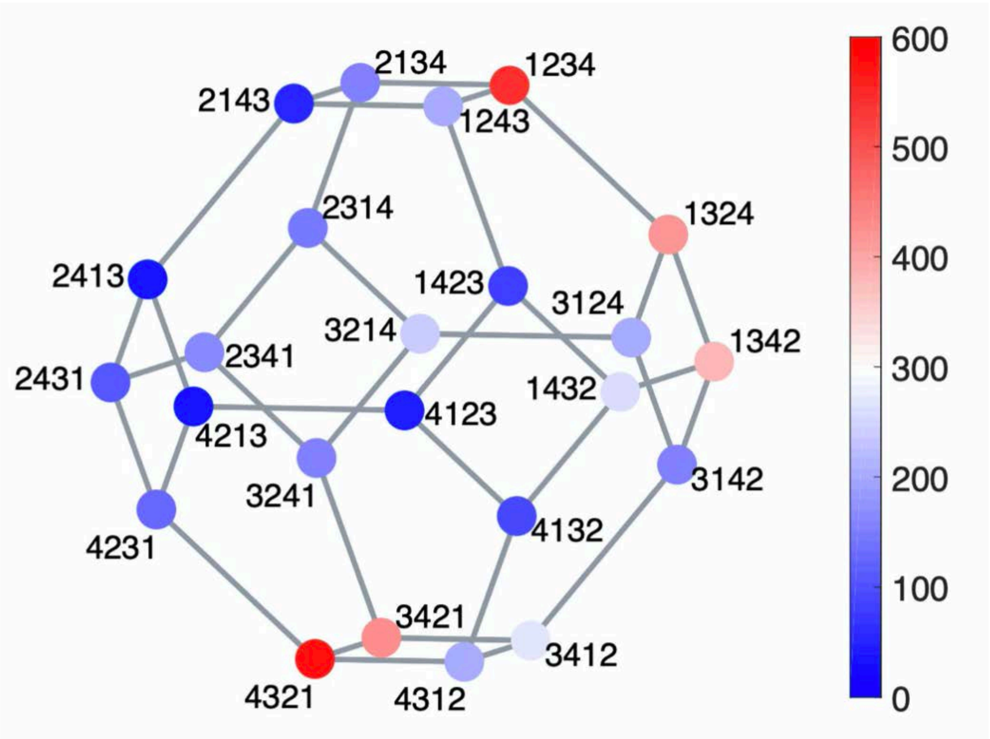
- Background on ranked choice voting, signals on the permutahedron, and representation theory
- Design of tight spectral frames for ranked data analysis
- **Illustrative analysis examples**
- Computational challenges and efficient algorithms
- Extensions and ongoing work

Minneapolis Election Data Revisited

Four Candidates:

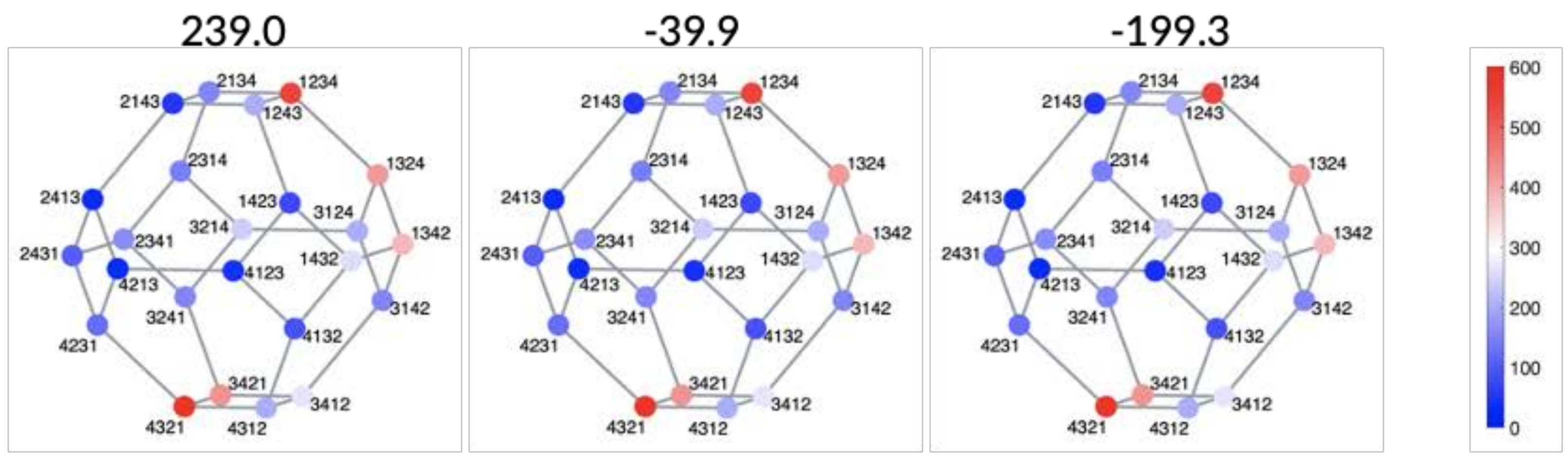
1. Ginger Jentzen
   (Socialist-Alternative)
2. Samantha Pree-Stinson
   (Green)
3. Steven Fletcher
   (Democratic-Farmer-Labor, elected)
4. Tim Bildsoe
   (Democratic-Farmer-Labor)

1st	2nd	3rd	4th	Count
4	3	2	1	574
4	3	1	2	201
4	2	3	1	131
4	2	1	3	32
4	1	3	2	89
4	1	2	3	46
3	4	2	1	422
3	4	1	2	271
3	2	4	1	159
3	2	1	4	243
3	1	4	2	156
3	1	2	4	204
2	4	3	1	111
2	4	1	3	30
2	3	4	1	161
2	3	1	4	145
2	1	4	3	56
2	1	3	4	153
1	4	3	2	255
1	4	2	3	77
1	3	4	2	376
1	3	2	4	421
1	2	4	3	204
1	2	3	4	538

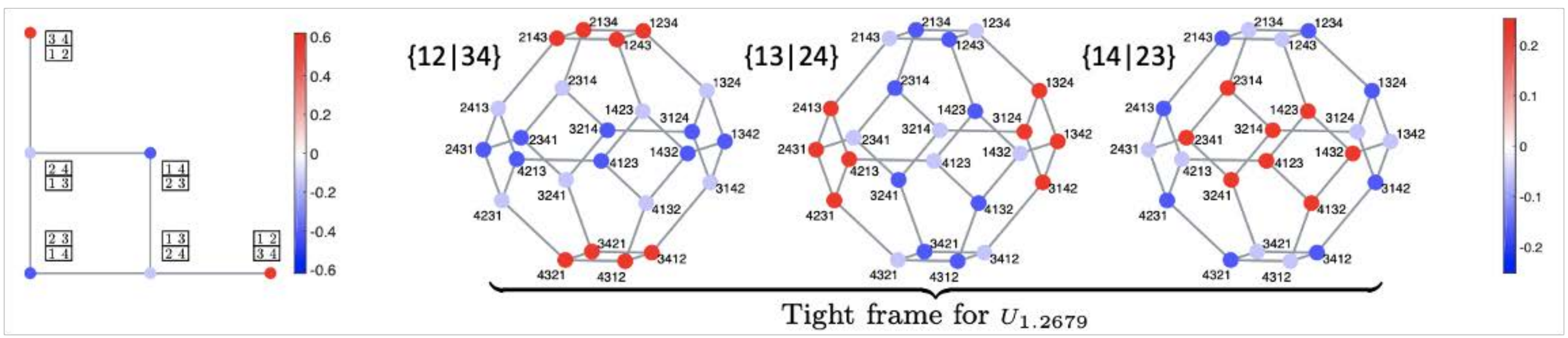


Candidate	First Choice	Second Choice	Third Choice	Fourth Choice
Ginger Jentzen	1871	704	922	1558
Samantha Pree-Stinson	656	1307	1744	1348
Steve Fletcher	1455	1878	1277	445
Tim Bildsoe	1073	1166	1112	1704

Analysis Example
Inner Product Between Signal and Each Atom



Signal (shown three times)



Frame vectors

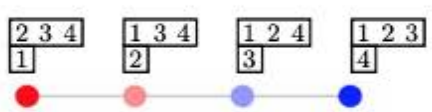
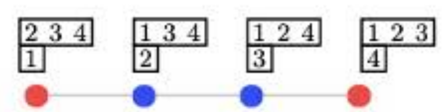
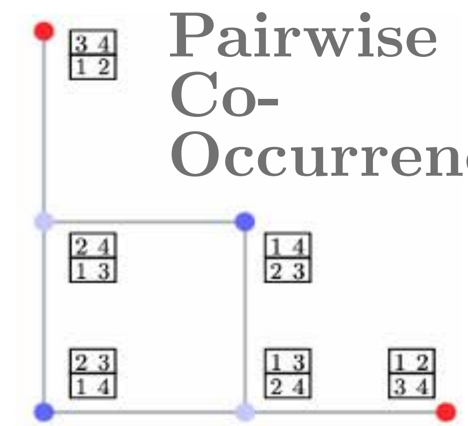
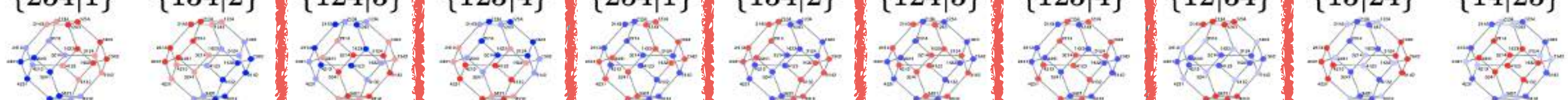
- Large positive coefficient: designated pairs of candidates appear together at the beginning or end of the rankings; often indicates similar politically
- Large negative coefficient: for each pair, often the case that one appears in the first two slots and the other appears in the last two slots
- Can also view the inner product as projecting the signal down from the permutahedron to the Schreier graph according to the characteristic matrix, and then taking the inner product there

Interpretation of Specific Analysis Coefficients

Energy Decomposition by Shape-Eigenvalue Pair

γ	(4)	$(3,1)$			$(2,2)$		$(2,1,1)$			$(1,1,1,1)$
λ	0	0.586	2	3.414	1.268	4.732	2.586	4	5.414	6
$\sum_{\bar{\pi}} \lvert\langle \mathbf{g}, \varphi_{\gamma,\lambda,\bar{\pi}} \rangle\rvert^2$	1064709.4	147617.5	192845.1	14739.0	98412.8	39162.5	13878.0	32979.6	1085.0	1820.0

Specific Analysis Coefficients

γ	$(3,1)$								$(2,2)$		
λ	0.586				2				1.268		
$\mathbf{v}_\lambda$	**Individual Popularity**  Positive: Popular · Negative: Unpopular				**Polarization**  Positive: Polarizing · Negative: Ranked in middle				**Pairwise Co-Occurrence** 		
$\bar{\pi}$	{234|1}	{134|2}	{124|3}	{123|4}	{234|1}	{134|2}	{124|3}	{123|4}	{12|34}	{13|24}	{14|23}
$\varphi_{\gamma,\lambda,\bar{\pi}}$											
$\langle \mathbf{g}, \varphi_{\gamma,\lambda,\bar{\pi}} \rangle$	51.4	-201.6	290.8	-140.6	318.7	-185.1	-221.9	88.2	239.0	-39.9	-199.3

Example: Sushi Preference Data

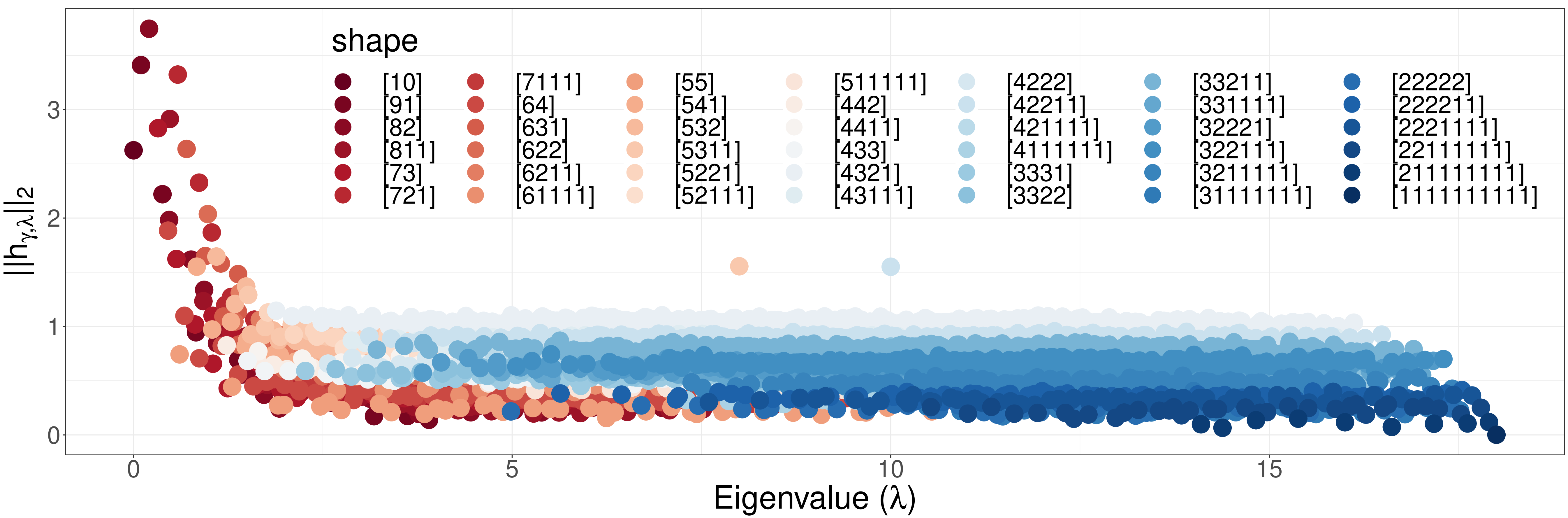
Index	Sushi Type
1	Shrimp
2	Sea eel
3	Tuna
4	Squid
5	Sea urchin
6	Salmon roe
7	Egg
8	Fatty tuna
9	Tuna roll
10 (0)	Cucumber roll

- $10! = 3.6$M permutations, 25.2M frame vectors
- Projections onto low-dimensional graphs provide insights

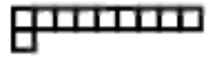
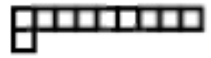
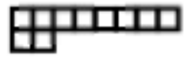
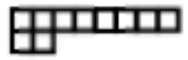
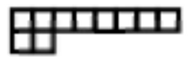
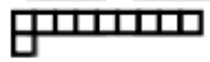
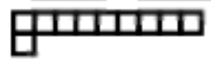
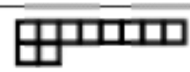
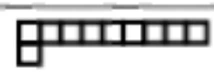
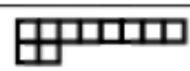
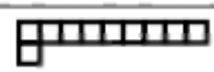
Kamishima, "Nantonac collaborative filtering: Recommendation based on order responses," SIGKDD, 2003

Energy Decomposition Across Shape-Eigenvalue Pairs

Index	Sushi Type
1	Shrimp
2	Sea eel
3	Tuna
4	Squid
5	Sea urchin
6	Salmon roe
7	Egg
8	Fatty tuna
9	Tuna roll
10 (0)	Cucumber roll



Analysis Coefficients with the Largest Magnitudes

γ	$\bar{\pi}$	λ	$\langle \mathbf{h}, \boldsymbol{\varphi}_{\gamma,\lambda,\bar{\pi}} \rangle$	$\lvert\langle \mathbf{h}, \boldsymbol{\varphi}_{\gamma,\lambda,\bar{\pi}} \rangle\rvert^2$
	$\{1234567890\}$	0	2.6248	6.8893
	$\{123456789\|0\}$	0.0979	-2.1513	4.6280
	$\{123456790\|8\}$	0.0979	1.9978	3.9912
	$\{12345679\|80\}$	0.2047	-1.7150	2.9413
	$\{12345689\|70\}$	0.2047	1.6543	2.7369
	$\{12345679\|8\|0\}$	0.4799	1.3471	1.8147
	$\{12456790\|38\}$	0.2047	1.3304	1.7699
	$\{123456890\|7\}$	0.0979	-1.1896	1.4150
	$\{123456780\|9\}$	0.3820	-1.1006	1.2112
	$\{1234569\|780\}$	0.3227	1.0659	1.1362
	$\{12345690\|78\}$	0.2047	-1.0400	1.0817
	$\{123456790\|8\}$	0.3820	1.0392	1.0800
	$\{12345689\|70\}$	0.4700	-1.0046	1.0093
	$\{123467890\|5\}$	0.3820	0.9604	0.9223

The most informative frame vectors often come from the earlier shapes and lower eigenvalues, corresponding to atoms that are smoother and more interpretable

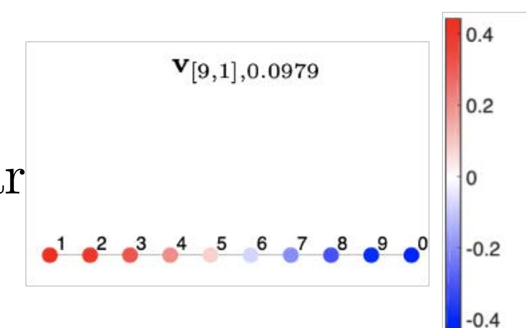
Interpretation of Analysis Coefficients

Index	Sushi Type
1	Shrimp
2	Sea eel
3	Tuna
4	Squid
5	Sea urchin
6	Salmon roe
7	Egg
8	Fatty tuna
9	Tuna roll
10 (0)	Cucumber roll

γ	t	λ	$\langle \mathbf{h}, \varphi_{\gamma,1,t} \rangle$	$\lvert\langle \mathbf{h}, \varphi_{\gamma,1,t} \rangle\rvert^2$
	{1234567890}	0	2.6248	6.8893
	{123456789|0}	0.0979	-2.1513	4.6280
	{123456790|8}	0.0979	1.9978	3.9912
	{12345679|80}	0.2047	-1.7150	2.9413
	{12345689|70}	0.2047	1.6543	2.7369
	{12345679|8|0}	0.4799	1.3471	1.8147
	{12456790|38}	0.2047	1.3304	1.7699
	{123456890|7}	0.0979	-1.1896	1.4150
	{123456780|9}	0.3820	-1.1006	1.2112
	{1234569|780}	0.3227	1.0659	1.1362
	{12345690|78}	0.2047	-1.0400	1.0817
	{123456790|8}	0.3820	1.0392	1.0800
	{12345689|70}	0.4700	-1.0046	1.0093
	{123467890|5}	0.3820	0.9604	0.9223

Most Popular?

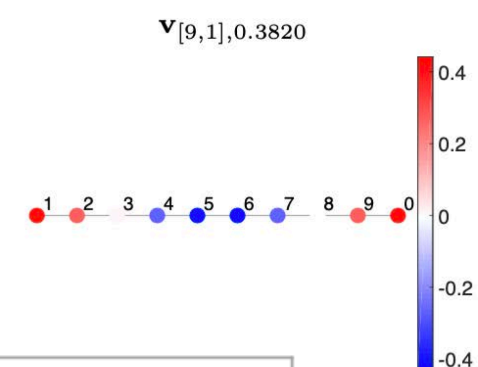
- Positive: popular
- Negative: unpopular



Candidate	Coefficient
10 (Cucumber)	-2.15
8 (Fatty Tuna)	1.99
7 (Egg)	-1.19

Most Polarizing?

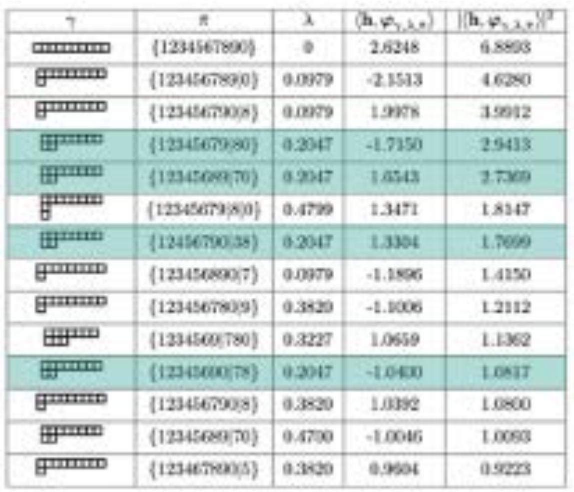
- Positive: polarizing
- Negative: commonly ranked in the middle



Candidate	Coefficient
9 (Tuna Roll)	-1.1006
8 (Fatty Tuna)	1.0392
5 (Sea Urchin)	0.9604

Interpretation of Analysis Coefficients

Index	Sushi Type
1	Shrimp
2	Sea eel
3	Tuna
4	Squid
5	Sea urchin
6	Salmon roe
7	Egg
8	Fatty tuna
9	Tuna roll
10 (0)	Cucumber roll



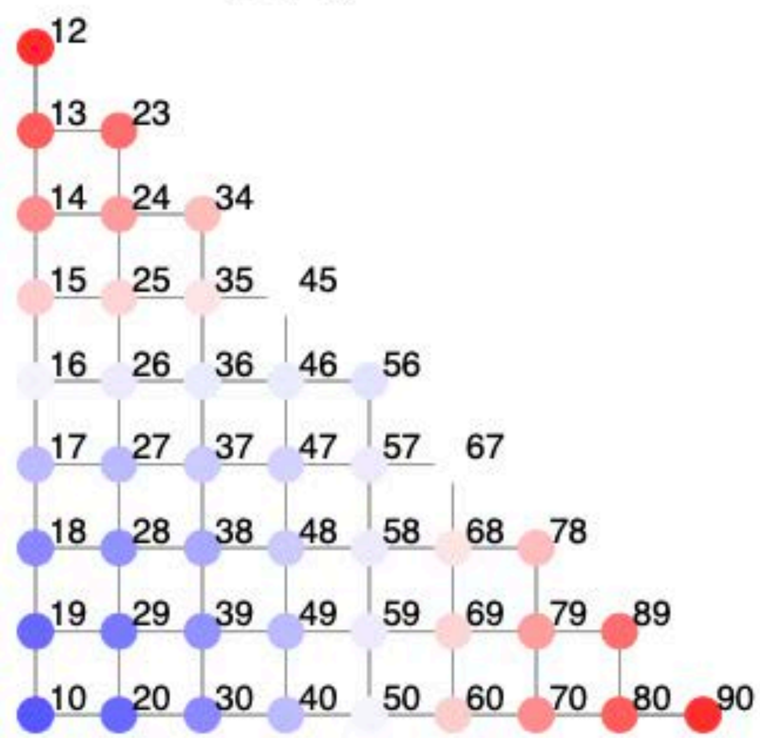
γ	π	λ	$\langle \mathbf{h}, \varphi_{\gamma,\lambda,\pi} \rangle$	$\lvert\langle \mathbf{h}, \varphi_{\gamma,\lambda,\pi} \rangle\rvert^2$
	{1234567890}	0	2.6248	6.8893
	{123456789|0}	0.0979	-2.1513	4.6280
	{123456790|8}	0.0979	1.9978	3.9912
	{12345679|80}	0.2047	-1.7150	2.9413
	{12345689|70}	0.2047	1.6543	2.7369
	{12345679|8|0}	0.4799	1.3471	1.8147
	{12456790|38}	0.2047	1.3304	1.7699
	{123456890|7}	0.0979	-1.1896	1.4150
	{123456780|9}	0.3820	-1.1006	1.2112
	{1234569|780}	0.3227	1.0659	1.1362
	{12345690|78}	0.2047	-1.0400	1.0817
	{123456790|8}	0.3820	1.0392	1.0800
	{12345689|70}	0.4700	-1.0046	1.0093
	{123467890|5}	0.3820	0.9604	0.9223

Pairwise Co-Occurence?

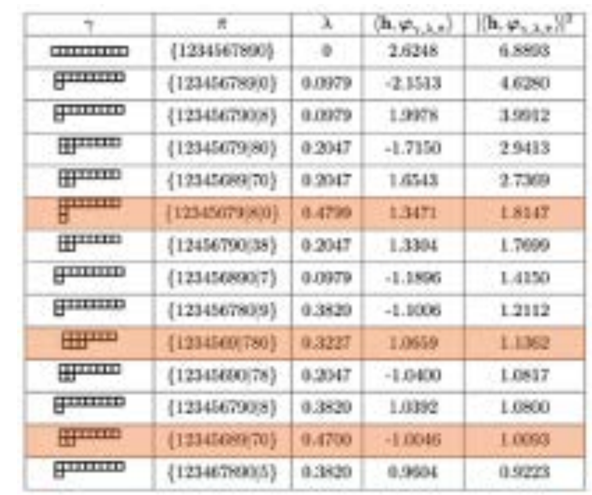
- Positive: ranked together
- Negative: ranked far apart

Candidates	Coefficient
8 (Fatty Tuna), 10 (Cucumber)	-1.7150
7 (Egg), 10 (Cucumber)	1.6543
3 (Tuna), 8 (Fatty Tuna)	1.3304
7 (Egg), 8 (Fatty Tuna)	-1.0400

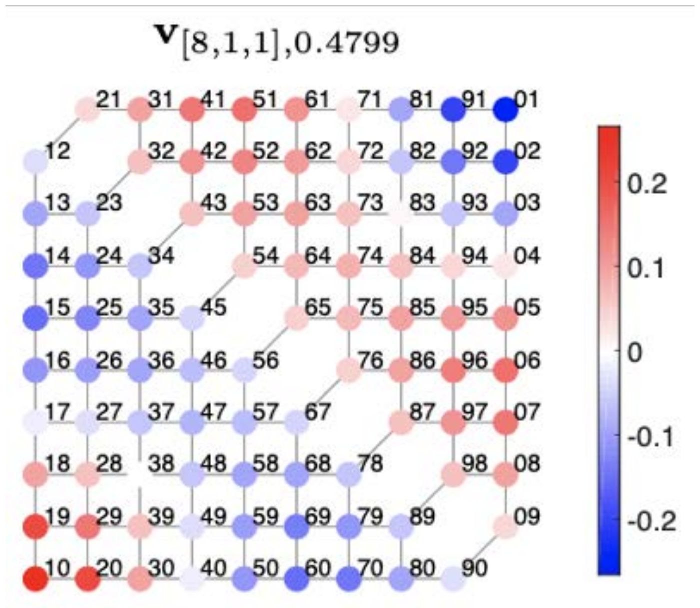
$\mathbf{v}_{[8,2],0.2047}$



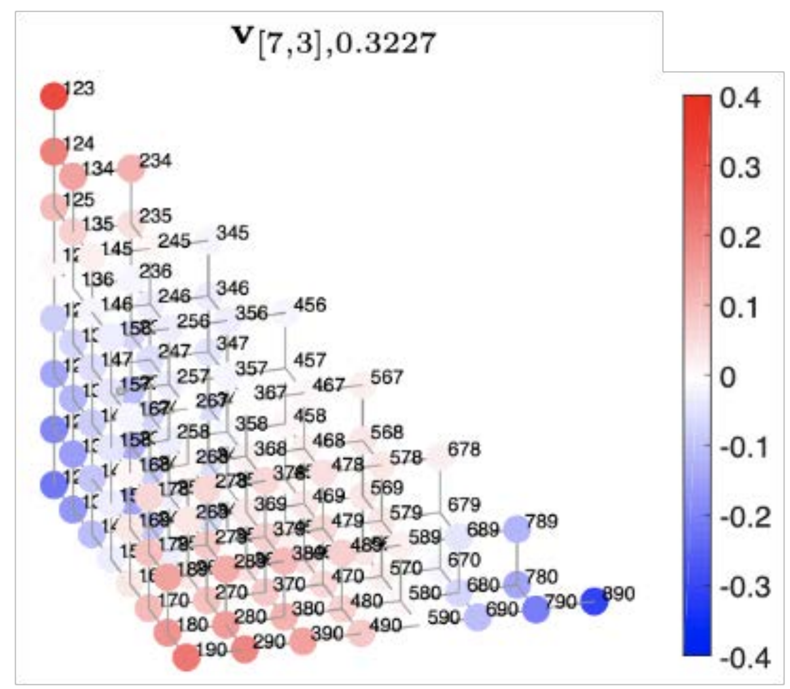
Interpretation of Analysis Coefficients



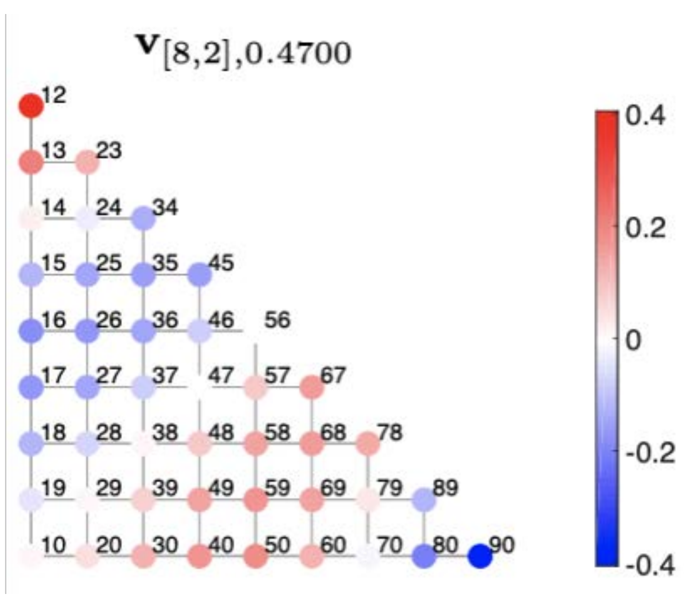
γ	π	λ	$\langle \mathbf{h}, \varphi_{\gamma,\lambda,\pi} \rangle$	$\lvert\langle \mathbf{h}, \varphi_{\gamma,\lambda,\pi} \rangle\rvert^2$
	{1234567890}	0	2.6248	6.8893
	{123456789|0}	0.0979	-2.1513	4.6280
	{123456790|8}	0.0979	1.9978	3.9912
	{12345679|80}	0.2047	-1.7150	2.9413
	{12345689|70}	0.2047	1.6543	2.7369
	{12345679|8|0}	0.4799	1.3471	1.8147
	{12456790|38}	0.2047	1.3304	1.7699
	{123456890|7}	0.0979	-1.1896	1.4150
	{123456780|9}	0.3820	-1.1006	1.2112
	{1234569|780}	0.3227	1.0659	1.1362
	{12345690|78}	0.2047	-1.0400	1.0817
	{123456790|8}	0.3820	1.0392	1.0800
	{12345689|70}	0.4700	-1.0046	1.0093
	{123467890|5}	0.3820	0.9604	0.9223



Fatty tuna and egg:
Fatty tuna ranked high, egg ranked low



Fatty tuna, egg, and cucumber roll:
Two ranked low, one ranked high



Egg and cucumber roll (negative):
Both ranked very low

Outline

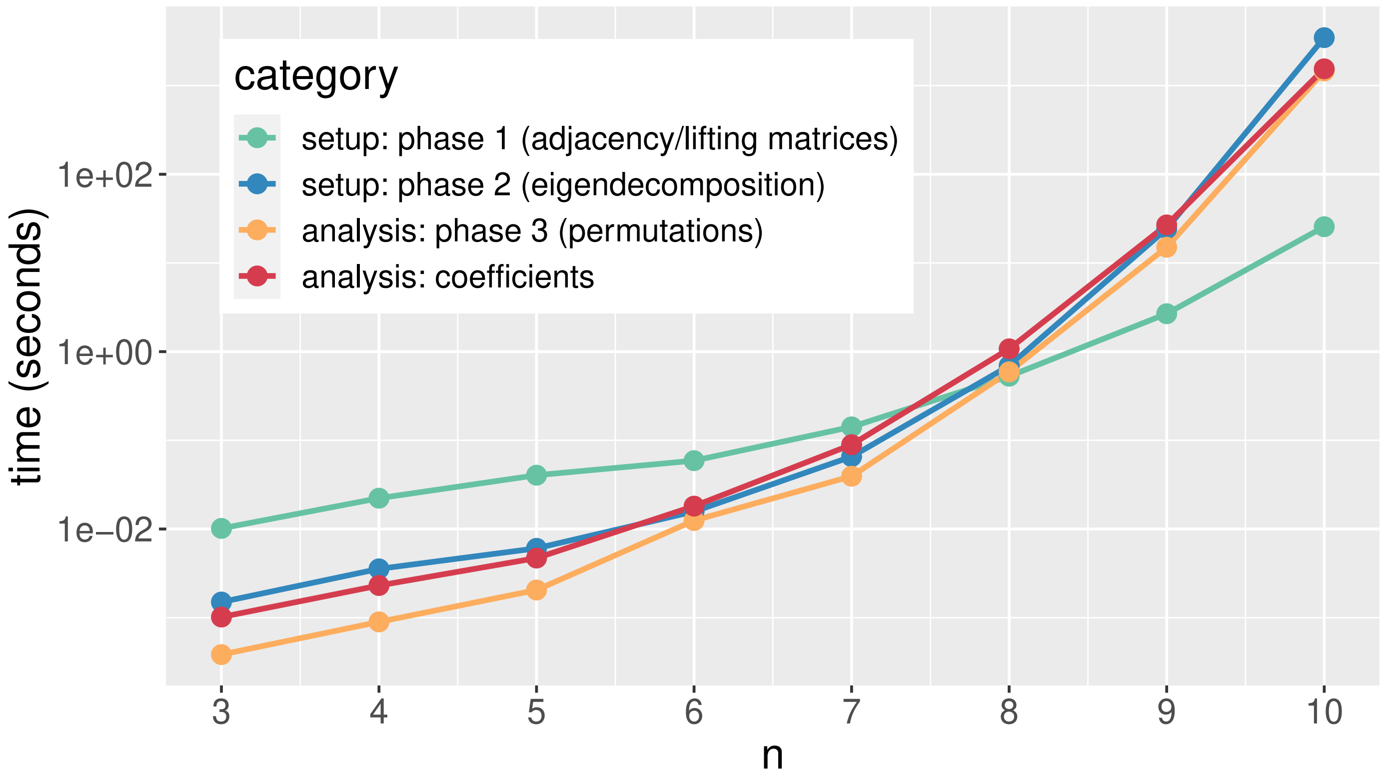
- Background on ranked choice voting, signals on the permutahedron, and representation theory
- Design of tight spectral frames for ranked data analysis
- Illustrative analysis examples
- **Computational challenges and efficient algorithms**
- Extensions and ongoing work

Computational Challenges

- Challenges numerous:
  - Not tractable to naively compute permutahedron eigenvectors with more than 7 or 8 candidates
  - Not even tractable to naively compute the Laplacian eigenvectors of the Schreier graphs
  - Can't store all dictionary atoms naively - memory issues
  - Number of atoms grows faster than $n!$; i.e., redundancy increases as n increases
- Need specialized algorithms that take advantage of the symmetry and structure present in the permutahedron
- The development of these algorithms actually led us to many new and interesting theoretical questions

Computational Strategies & Times

- Reduce the number of atoms by focusing on certain eigen-symmetry subspaces
- Include as much of the computation as possible into a setup portion that only has to be done once for each n
- Cleverly leverage recursion, symmetry, deflation, and dimension reduction!



Outline

- Background on ranked choice voting, signals on the permutahedron, and representation theory
- Design of tight spectral frames for ranked data analysis
- Illustrative analysis examples
- Computational challenges and efficient algorithms
- **Extensions and ongoing work**

Ongoing Work

- Investigate specific application questions
  - How to do polling for ranked choice voting?
  - How to detect voting fraud / anomalies?
  - How to track multiple objects in computer vision, when there is a spacial component to the tracks?
  - Connections with rank aggregation methods?
- Closed form computation of the Schreier eigenvectors and eigenvalues
- Approximations to speed up computations

Extension: Tight Frames for Analyzing Data on Other Cayley Graphs, Groups, and Combinatorial Structures

- The machinery we developed can be extended to construct tight spectral frames on any finite group
- Example: Applied to the hyperoctohedral group, we can analyze data on signed permutations
- Frame vectors still lifted from Schreier coset graphs (quotient graphs)
- Utilizes double-centralizer (Schur-Weyl) duality between the left group action and the right group action

Extension: Partial and Incomplete Rankings

- Partial rankings: allow ties
  - Example: rank top 3 and assume the others are tied at the end
- Incomplete rankings: voters only rank a subset of the candidates (saying nothing about the others)
  - Related to balanced incomplete block design
- One method is to split (equally) each vote onto the vertices of the permutahedron consistent with that vote and then use the proposed tight frame
- However, more nuanced analyses might be possible, by, e.g., constructing a different graph to capture all possible partial/incomplete rankings

Key Takeaways

- Permutahedron graph captures an appropriate notion of distance for ranked choice voting data
- Interesting connections between graph signal processing and combinatorial representation theory
- Developed a method to construct tight frames for ranked data consisting of atoms that are contained in eigen-shape spaces; i.e., they have interpretable smoothness and symmetry properties
- Can leverage the resulting transform to identify polarizing candidates, perform party detection / candidate clustering, etc.