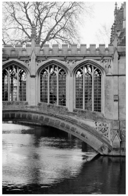
Part I: Rank Aggregation via Hodge Theory

Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, Yinyu Ye

NIPS Workshop on Advances in Ranking

December 11, 2009



Old and New Problems with Rank Aggregation

- Old Problems
  - **Condorcet's paradox:** majority vote intransitive $a \succeq b \succeq c \succeq a$. [Condorcet, 1785]
  - **Arrow's impossibility:** any sufficiently sophisticated preference aggregation must exhibit intransitivity. [Arrow, 1950], [Sen, 1970]
  - **Kemeny optimal is NP-hard:** even with just 4 voters. [Dwork-Kumar-Naor-Sivakumar, 2001]
  - **Empirical studies:** lack of majority consensus common in group decision making.
- New Problems
  - **Incomplete data:** typically about 1%.
  - **Imbalanced data:** power-law, heavy-tail distributed votes.
  - **Cardinal data:** given in terms of scores or stochastic choices.
  - **Voters' bias:** extreme scores, no low scores, no high scores.

Pairwise Ranking as a Solution

Example (Netflix Customer-Product Rating)

- 480189-by-17770 customer-product rating matrix A.
- **incomplete**: 98.82% of values missing.
- **imbalanced**: number of ratings on movies varies from 10 to 220,000.

- **Incompleteness:** pairwise comparison matrix X almost complete! 0.22% of the values are missing.
- **Intransitivity:** define model based on minimizing this as objective.
- **Cardinal:** use this to our advantage; linear regression instead of order statistics.
- **Complexity:** numerical linear algebra instead of combinatorial optimization.
- **Imbalance:** use this to choose an inner product/metric.
- **Bias:** pairwise comparisons alleviate this.

What We Seek

Ordinal: Intransitivity, $a \succeq b \succeq c \succeq a$.

Cardinal: Inconsistency, $X_{ab} + X_{bc} + X_{ca} \neq 0$.

- Want global ranking of the alternatives if a reasonable one exists.
- Want certificate of reliability to quantify validity of global ranking.
- If no meaningful global ranking, analyze nature of inconsistencies.

A basic tenet of data analysis is this: If you've found some structure, take it out, and look at what's left. Thus to look at second order statistics it is natural to subtract away the observed first order structure. This leads to a natural decomposition of the original data into orthogonal pieces.

Persi Diaconis, 1987 Wald Memorial Lectures

Orthogonal Pieces of Ranking

- Hodge decomposition:

$$\textit{aggregate pairwise ranking} = \textit{consistent} \oplus \textit{locally inconsistent} \oplus \textit{globally inconsistent}$$

- Consistent component gives global ranking.
- Total size of inconsistent components gives certificate of reliability.
- Local and global inconsistent components can do more than just certifying the global ranking.

Analyzing Inconsistencies

- Locally inconsistent rankings should be acceptable.
  - Inconsistencies in items ranked closed together but not in items ranked far apart.
  - Ordering of 4th, 5th, 6th ranked items cannot be trusted but ordering of 4th, 50th, 600th ranked items can.
  - E.g. no consensus for hamburgers, hot dogs, pizzas, and no consensus for caviar, foie gras, truffle, but clear preference for latter group.
- Globally inconsistent rankings ought to be rare.

Theorem (Kahle, 2007)

Erdős-Rényi $G(n, p)$, n alternatives, comparisons occur with probability p, clique complex χ_G almost always have zero 1-homology, unless

$$\frac{1}{n^2} \ll p \ll \frac{1}{n}.$$

Basic Model

- Ranking data live on **pairwise comparison graph** $G = (V, E)$; V: set of alternatives, E: pairs of alternatives to be compared.
- Optimize over model class $\mathcal{M}$

$$\min_{X \in \mathcal{M}} \sum\nolimits_{\alpha,i,j} w_{ij}^{\alpha}(X_{ij} - Y_{ij}^{\alpha})^2.$$

- Y_{ij}^{α} measures preference of i over j of voter α. Y^{α} skew-symmetric.
- w_{ij}^{α} metric; 1 if α made comparison for $\{i, j\}$, 0 otherwise.
- Kemeny optimization:

$$\mathcal{M}_K = \{X \in \mathbb{R}^{n \times n} \mid X_{ij} = \operatorname{sign}(s_j - s_i),\ s : V \to \mathbb{R}\}.$$

- Relaxed version:

$$\mathcal{M}_G = \{X \in \mathbb{R}^{n \times n} \mid X_{ij} = s_j - s_i,\ s : V \to \mathbb{R}\}.$$

- Rank-constrained least squares with skew-symmetric matrix variables.

Rank Aggregation

- Previous problem may be reformulated

$$\min_{X \in \mathcal{M}_G} \|X - \bar{Y}\|_{F,w}^2 = \min_{X \in \mathcal{M}_G} \left[\sum_{\{i,j\} \in E} w_{ij} (X_{ij} - \bar{Y}_{ij})^2 \right]$$

  where

$$w_{ij} = \sum_\alpha w_{ij}^\alpha \quad \text{and} \quad \bar{Y}_{ij} = \sum_\alpha w_{ij}^\alpha Y_{ij}^\alpha \, / \, \sum_\alpha w_{ij}^\alpha .$$

- Why not just aggregate over scores directly? Mean score is a **first order** statistics and is inadequate because
  - most voters would rate just a very small portion of the alternatives,
  - different alternatives may have different voters, mean scores affected by individual rating scales.
- Use higher order statistics.

Formation of Pairwise Ranking

Linear Model: average score difference between i and j over all who have rated both,

$$Y_{ij} = \frac{\sum_k (X_{kj} - X_{ki})}{\#\{k \mid X_{ki}, X_{kj} \text{ exist}\}}.$$

Log-linear Model: logarithmic average score ratio of positive scores,

$$Y_{ij} = \frac{\sum_k (\log X_{kj} - \log X_{ki})}{\#\{k \mid X_{ki}, X_{kj} \text{ exist}\}}.$$

Linear Probability Model: probability j preferred to i in excess of purely random choice,

$$Y_{ij} = \Pr\{k \mid X_{kj} > X_{ki}\} - \frac{1}{2}.$$

Bradley-Terry Model: logarithmic odd ratio (logit),

$$Y_{ij} = \log \frac{\Pr\{k \mid X_{kj} > X_{ki}\}}{\Pr\{k \mid X_{kj} < X_{ki}\}}.$$

Functions on Graph

$G = (V, E)$ undirected graph. V vertices, $E \in \binom{V}{2}$ edges, $T \in \binom{V}{3}$ triangles/3-cliques. $\{i, j, k\} \in T$ iff $\{i, j\}, \{j, k\}, \{k, i\} \in E$.

- **Function on vertices:** $s : V \to \mathbb{R}$
- **Edge flows:** $X : V \times V \to \mathbb{R}$, $X(i, j) = 0$ if $\{i, j\} \notin E$,

$$X(i, j) = -X(j, i) \quad \text{for all } i, j.$$

- **Triangular flows:** $\Phi : V \times V \times V \to \mathbb{R}$, $\Phi(i, j, k) = 0$ if $\{i, j, k\} \notin T$,

$$\begin{aligned}\Phi(i, j, k) &= \Phi(j, k, i) = \Phi(k, i, j) \\ &= -\Phi(j, i, k) = -\Phi(i, k, j) = -\Phi(k, j, i) \quad \text{for all } i, j, k.\end{aligned}$$

- Physics: s, X, Φ potential, alternating vector/tensor field.
- Topology: s, X, Φ 0-, 1-, 2-cochain.
- Ranking: s scores/utility, X pairwise rankings, Φ triplewise rankings

Operators

- **Gradient:** $\text{grad} : L^2(V) \to L^2(E)$,
$$(\text{grad}\, s)(i,j) = s_j - s_i.$$
- **Curl:** $\text{curl} : L^2(E) \to L^2(T)$,
$$(\text{curl}\, X)(i,j,k) = X_{ij} + X_{jk} + X_{ki}.$$
- **Divergence:** $\text{div} : L^2(E) \to L^2(V)$,
$$(\text{div}\, X)(i) = \sum\nolimits_j w_{ij} X_{ij}.$$
- **Graph Laplacian:** $\Delta_0 : L^2(V) \to L^2(V)$,
$$\Delta_0 = \text{div} \circ \text{grad}\,.$$
- **Graph Helmholtzian:** $\Delta_1 : L^2(E) \to L^2(E)$,
$$\Delta_1 = \text{curl}^* \circ \text{curl} - \text{grad} \circ \text{div}\,.$$

Some Properties

- im(grad): pairwise rankings that are gradient of score functions, i.e. consistent or *integrable*.
- ker(div): $\operatorname{div} X(i)$ measures the inflow-outflow sum at i; $\operatorname{div} X(i) = 0$ implies alternative i is preference-neutral in all pairwise comparisons; i.e. inconsistent rankings of the form $a \succeq b \succeq c \succeq \cdots \succeq a$.
- ker(curl): pairwise rankings with zero flow-sum along any triangle.
- $\ker(\Delta_1) = \ker(\mathrm{curl}) \cap \ker(\mathrm{div})$: globally inconsistent or *harmonic* rankings; no inconsistencies due to small loops of length 3, i.e. $a \succeq b \succeq c \succeq a$, but inconsistencies along larger loops of lengths > 3.
- $\mathrm{im}(\mathrm{curl}^*)$: locally inconsistent rankings; non-zero curls along triangles.
- $\operatorname{div} \circ \operatorname{grad}$ is vertex Laplacian, $\mathrm{curl} \circ \mathrm{curl}^*$ is edge Laplacian.

Boundary of a Boundary is Empty

Algebraic topology in a slogan: (co)boundary of (co)boundary is null.

$$\text{Global} \xrightarrow{\text{grad}} \text{Pairwise} \xrightarrow{\text{curl}} \text{Triplewise}$$

and so

$$\text{Global} \xleftarrow{\text{grad}^*(=: -\text{div})} \text{Pairwise} \xleftarrow{\text{curl}^*} \text{Triplewise}.$$

We have

$$\text{curl} \circ \text{grad} = 0, \qquad \text{div} \circ \text{curl}^* = 0.$$

This implies

- global rankings are transitive/consistent,
- no need to consider rankings beyond triplewise.

You've probably seen Hodge theory

$$V = \{F : \mathbb{R}^3 \backslash X \to \mathbb{R}^3 \mid \nabla \times F = 0\}; \quad W = \{F = \nabla g\}; \quad \dim(V/W) = ?$$

Helmholtz/Hodge Decomposition

- **Vector calculus:** vector field $\mathbf{F}$ resolvable into irrotational (curl-free), solenoidal (divergence-free), harmonic parts, $\mathbf{F} = -\nabla\varphi + \nabla \times \mathbf{A} + H$ where φ scalar potential, $\mathbf{A}$ vector potential.
- **Linear algebra:** every skew-symmetric matrix X can be written as sum of three skew-symmetric matrices $X = X_1 + X_2 + X_3$ where $X_1 = se^\top - es^\top$, $X_2(i,j) + X_2(j,k) + X_2(k,i) = 0$.
- **Graph theory:** orthogonal decomposition of network flows into acyclic and cyclic components.

Theorem (Helmholtz decomposition)

$G = (V, E)$ undirected, unweighted graph. Δ_1 its Helmholtzian. The space of edge flows admits orthogonal decomposition

$$L^2(E) = \operatorname{im}(\operatorname{grad}) \oplus \ker(\Delta_1) \oplus \operatorname{im}(\operatorname{curl}^*).$$

Furthermore, $\ker(\Delta_1) = \ker(\delta_1) \cap \ker(\delta_0^*) = \ker(\operatorname{curl}) \cap \ker(\operatorname{div})$.

Rank Aggregation Problem Revisited

Recall our formulation

$$\min_{X \in \mathcal{M}_G} \|X - \bar{Y}\|_{2,w}^2 = \min_{X \in \mathcal{M}_G} \left[\sum_{\{i,j\} \in E} w_{ij}(X_{ij} - \bar{Y}_{ij})^2 \right].$$

The exact case is:

Problem (Integrability of Vector Fields)

Does there exist a global ranking function, $s : V \to \mathbb{R}$, such that

$$X_{ij} = s_j - s_i =: (\operatorname{grad} s)(i,j)?$$

Answer: There are non-integrable vector fields, i.e.

$$V = \{F : \mathbb{R}^3 \backslash X \to \mathbb{R}^3 \mid \nabla \times F = 0\}; \quad W = \{F = \nabla g\}; \quad \dim(V/W) > 0.$$

Harmonic Rankings?

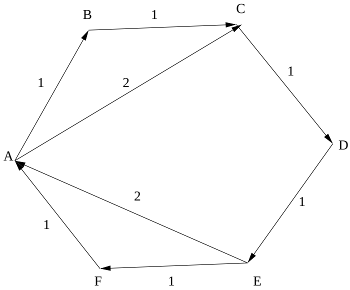


Figure: Locally consistent but globally inconsistent harmonic ranking.

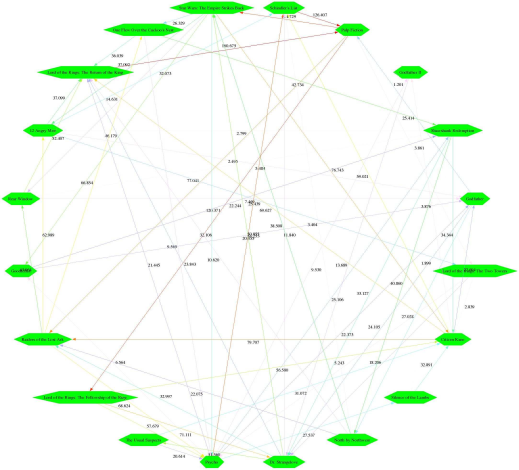


Figure: Harmonic ranking from a truncated Netflix movie-movie network

College Ranking

	Kendall τ-distance								
	RAE'01	in-degree	out-degree	HITS authority	HITS hub	PageRank	Hodge ($k=1$)	Hodge ($k=2$)	Hodge ($k=4$)
RAE'01	0	0.0994	0.1166	0.0961	0.1115	0.0969	0.1358	0.0975	0.0971
in-degree	0.0994	0	0.0652	0.0142	0.0627	0.0068	0.0711	0.0074	0.0065
out-degree	0.1166	0.0652	0	0.0672	0.0148	0.0647	0.1183	0.0639	0.0647
HITS authority	0.0961	0.0142	0.0672	0	0.0627	0.0119	0.0736	0.0133	0.0120
HITS hub	0.1115	0.0627	0.0148	0.0627	0	0.0615	0.1121	0.0607	0.0615
PageRank	0.0969	0.0068	0.0647	0.0119	0.0615	0	0.0710	0.0029	0.0005
Hodge ($k=1$)	0.1358	0.0711	0.1183	0.0736	0.1121	0.0710	0	0.0692	0.0709
Hodge ($k=2$)	0.0975	0.0074	0.0639	0.0133	0.0607	0.0029	0.0692	0	0.0025
Hodge ($k=3$)	0.0971	0.0065	0.0647	0.0120	0.0615	0.0005	0.0709	0.0025	0

Table: Kendall τ-distance between different global rankings. Note that HITS authority gives the nearest global ranking to the research score RAE'01, while Hodge decompositions for $k \geq 2$ give closer results to PageRank which is the second closest to the RAE'01.