Combinatorial Hodge Theory and a Geometric Approach to Ranking

Yuan Yao

2008 SIAM Annual Meeting

San Diego, July 7, 2008

with Lek-Heng Lim et al.
1 Ranking on networks (graphs)
   - Netflix example
   - Skew-symmetric matrices and network flows

2 Combinatorial Hodge Theory
   - Discrete Differential Geometry
   - Hodge Decomposition Theorem
   - Rank aggregation as a linear projection

3 Why pairwise works for Netflix: a spectral embedding view

4 Conclusions
Ranking on Networks (Graphs)

- “Multi-criteria” rank/decision systems
  - Amazon or Netflix’s recommendation system (user-product)
  - Interest ranking in social networks (person-interest)
  - Financial analyst recommendation system (analyst-stock)
  - Voting (voter-candidate)

- “Peer-review” systems
  - publication citation systems (paper-paper)
  - Google’s webpage ranking (web-web)
  - eBay’s reputation system (customer-customer)
Ranking data are

- incomplete: partial list or pairwise (e.g. \( \sim 1\% \) in Netflix)
- unbalanced: varied distributed votes (e.g. power law)
- cardinal: scores or stochastic choices

Implicitly or explicitly, ranking data may be viewed to live on a simple graph \( G = (V, E) \)

- \( V \): set of alternatives (products, interests,...) to be ranked
- \( E \): pairs of alternatives compared
Example: Netflix

Example (Netflix Customer-Product Rating)

- 480189-by-17770 customer-product 5-star rating matrix $X$
- $X$ is incomplete, with 98.82% of missing values

However,

- pairwise comparison graph $G = (V, E)$ is very dense!
- only 0.22% edges are missed, almost complete
- rank aggregation without estimating missing values!
- unbalanced: number of raters on $e \in E$ varies.

Caveat: we are not trying to solve the Netflix prize problem
The first-order statistics, mean score for each product, is often inadequate because of the following:

- most customers just rate a very small portion of the products
- different products might have different raters, whence mean scores involve noise due to arbitrary individual rating scales
- not able to characterize the inconsistency (ubiquitous for rank aggregation: Arrow’s impossibility theorem)

How about high order statistics?
From $1^{st}$ Order to $2^{nd}$ Order: Pairwise Rankings

- **Linear Model**: average score difference between product $i$ and $j$ over all customers who have rated both of them,

  $$w_{ij} = \frac{\sum_k (X_{kj} - X_{ki})}{\# \{k : X_{ki}, X_{kj} \text{ exist}\}},$$

  is translation invariant.

- **Log-linear Model**: when all the scores are positive, the logarithmic average score ratio,

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

  is invariant up to a multiplicative constant.
More Invariants

- **Linear Probability Model**: the probability that product \( j \) is preferred to \( i \) in excess of a purely random choice,

\[
    w_{ij} = \Pr\{k : X_{kj} > X_{ki}\} - \frac{1}{2}.
\]

This is invariant up to a monotone transformation.

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

\[
    w_{ij} = \log \frac{\Pr\{k : X_{kj} > X_{ki}\}}{\Pr\{k : X_{kj} < X_{ki}\}}.
\]

This is invariant up to a monotone transformation.
Skew-Symmetric Matrices of Pairwise Rankings

All such models induce (sparse) skew-symmetric matrices of $|V|\times|V|$ (or 2-alternating tensor),

$$w_{ij} = \begin{cases} -w_{ji}, & \{i,j\} \in E \\ NA, & \text{otherwise} \end{cases}$$

where $G = (V, E)$ is a pairwise comparison graph. 
Note: Such a skew-symmetric matrix induces a pairwise ranking network flow on graph $G$. 
Pairwise Ranking Flow for IMDB Top 20 Movies

Figure: Pairwise ranking flow on a complete graph
Rank Aggregation Problem

Hardness:

- Arrow-Sen’s impossibility theorems in social choice theory
- Kemeny-Snell optimal ordering is NP-hard to compute
- Spectral analysis on $S_n$ is impractical for large $n$ since $|S_n| = n!$

Our approach:

Problem (Rank Aggregation)

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

$$w_{ij} = v_j - v_i =: \delta_0(v)(i,j)?$$

Equivalently, does there exist a scalar field $v : V \rightarrow \mathbb{R}$ whose gradient field gives the flow $w$?
From multivariate calculus, there are non-integrable vector fields (cf. movie *A Beautiful Mind*). A combinatorial version:

Figure: No global ranking \( \nu \) gives \( w_{ij} = \nu_j - \nu_i \): (a) cyclic ranking, note \( w_{AB} + w_{BC} + w_{CA} \neq 0 \); (b) contains a 4-node cyclic flow \( A \rightarrow C \rightarrow D \rightarrow E \rightarrow A \), note on 3-clique \( \{A, B, C\} \) (also \( \{A, E, F\} \)), \( w_{AB} + w_{BC} + w_{CA} = 0 \)
Fact

For a skew-symmetric matrix $W = (w_{ij})$ associated with graph $G$, 

$$\exists v : w_{ij} = v_j - v_i \Rightarrow w_{ij} + w_{jk} + w_{ki} = 0, \forall 3\text{-clique } \{i, j, k\}$$

Note:

- **Transitivity subspace**: null triangular-trace or curl-free

$$\{W : w_{ij} + w_{jk} + w_{ki} = 0, \forall 3\text{-clique } \{i, j, k\}\}$$

Example in the last slide, (a) lies outside; (b) lies in this subspace, but not a gradient flow.
Hodge decomposition for skew-symmetric matrices

A skew-symmetric matrix $W$ associated with $G$ has an unique orthogonal decomposition

$$W = W_1 + W_2 + W_3$$

where

- $W_1$ satisfies (‘integrable’): $W_1(i,j) = v_j - v_i$ for some $v : V \to \mathbb{R}$;
- $W_2$ satisfies that
  - (‘curl-free’) $W_2(i,j) + W_2(j,k) + W_3(k,i) = 0$ for all 3-clique $\{i,j,k\}$
  - (‘divergence-free’) $\sum_j W_3(i,j) = 0$ for all edge $\{i,j\} \in E$;
- $W_3 \perp W_1$ and $W_3 \perp W_2$. 
Hodge decomposition for network flows

A network flows (e.g. pairwise rankings) on graph $G$ has an orthogonal decomposition into

$$
\text{gradient flow} + \text{locally acyclic (harmonic)} + \text{locally cyclic}
$$

where the first two components lie in the transitivity subspace and

- gradient flow is integrable to give a global ranking
- example (b) is locally acyclic, but cyclic on large scale (harmonic)
- example (a) is locally cyclic
We extend graph $G$ to a simplicial complex $\chi_G$ by attaching triangles

- 0-simplices $\chi^0_G$: vertices $V$
- 1-simplices $\chi^1_G$: edges $E$ such that comparison (i.e. pairwise ranking) between $i$ and $j$ exists
- 2-simplices $\chi^2_G$: triangles $\{i, j, k\}$ such that every edge exists

**Note:** it suffices here to construct $\chi_G$ up to dimension 2!

- global ranking $\nu : V \rightarrow \mathbb{R}$, 0-forms, vectors
- pairwise ranking $w(i, j) = -w(j, i)$ ($\forall (i, j) \in E$), 1-forms, skew-symmetric matrices, network flows
Space of $k$-Forms ($k$-cochains) and Metrics

- **$k$-forms:**

  \[ C^k(\chi_G, \mathbb{R}) = \{ u : \chi^{k+1}_G \to \mathbb{R}, u_{\sigma(i_0), \ldots, \sigma(i_k)} = \text{sign}(\sigma)u_{i_0, \ldots, i_k} \} \]

  for \((i_0, \ldots, i_k) \in \chi^{k+1}_G\), where \(\sigma \in S_{k+1}\) is a permutation on \((0, \ldots, k)\). Also \(k + 1\)-alternating tensors.

- One may associate \(C^k(\chi_G, \mathbb{R})\) with inner-products.

- In particular, the following inner-product on 1-forms is for unbalance issue in pairwise ranking

  \[ \langle w_{ij}, \omega_{ij} \rangle_D = \sum_{(i,j) \in E} D_{ij}w_{ij}\omega_{ij}, \quad w, \omega \in C^1(\chi_G, \mathbb{R}) \]

  where \(D_{ij} = |\{\text{customers who rate both } i \text{ and } j\}|\).
Discrete Exterior Derivatives (Coboundary Maps)

- **k-coboundary maps** $\delta_k : C^k(\chi_G, \mathbb{R}) \rightarrow C^{k+1}(\chi_G, \mathbb{R})$ are defined as the alternating difference operator

$$
(\delta_k u)(i_0, \ldots, i_{k+1}) = \sum_{j=0}^{k+1} (-1)^{j+1} u(i_0, \ldots, i_{j-1}, i_j, i_{j+1}, \ldots, i_{k+1})
$$

- $\delta_k$ plays the role of differentiation

- In particular,
  - $$(\delta_0 v)(i, j) = v_j - v_i =: (\text{grad} v)(i, j)$$
  - $$(\delta_1 w)(i, j, k) = (\pm) w_{ij} + w_{jk} + w_{ki} =: (\text{curl} w)(i, j, k)$$
    (triangular-trace of skew-symmetric matrix $(w_{ij})$)
Curl

**Definition**

For each triangle \( \{i, j, k\} \), the curl (triangular trace)

\[
\omega_{ij} + \omega_{jk} + \omega_{ki}
\]

measures the total flow-sum along the loop \( i \rightarrow j \rightarrow k \rightarrow i \).

- \((\delta_1 \omega)(i, j, k) = 0\) implies the flow is path-independent, which defines the triangular transitivity subspace.
Two directions of cochain maps:

- **Forward**
  \[ C^0 \xrightarrow{\delta_0} C^1 \xrightarrow{\delta_1} C^2, \]
  in other words,
  
  \[ \text{Global} \xrightarrow{\text{grad}} \text{Pairwise} \xrightarrow{\text{curl}} \text{Triplewise} \]

- **Backward**
  \[ C^2 \xleftarrow{\delta_0^* = \text{grad}^*} \xleftarrow{\delta_1^* = \text{curl}^*} 3\text{-alternating} \]
  
  - \( \text{grad}^* = \delta_0^* \) is the negative divergence
  - \( \text{curl}^* = \delta_1^* \), is the boundary operator, gives **triangular** (locally) **cyclic** pairwise rankings along triangles
Divergence

**Definition**

For each alternative $i \in V$, the divergence

$$(\text{div } w)(i) := -(\delta^T_0 w)(i) := \sum w_{i*}$$

measures the inflow-outflow sum at $i$.

- $(\delta^T_0 w)(i) = 0$ implies alternative $i$ is preference-neutral in all pairwise comparisons.
- divergence-free flow $\delta^T_0 w = 0$ is cyclic
- With metric $D$, conjugate operator gives weighted flow-sum

$$(\delta^*_0 w)(i) = \sum w_{ij}D_{ij} = (\delta^T_0 Dw)(i)$$
A Fundamental Property: Closed Map Property

- 'Boundary of boundary is empty': $\delta_{k+1} \circ \delta_k = 0$
- in particular
- • gradient flow is curl-free: $\text{curl} \circ \text{grad} = \delta_2 \circ \delta_1 = 0$
- • circular flow is divergence-free: $\text{div} \circ \text{curl}^* = \delta_1^* \circ \delta_2^* = 0$

This leads to the powerful combinatorial Laplacians.
Combinatorial Laplacians

Define the $k$-dimensional combinatorial Laplacian, $\Delta_k : C^k \rightarrow C^k$ by

$$\Delta_k = \delta_{k-1} \delta_{k-1}^* + \delta_k^* \delta_k, \quad k > 0$$

- $k = 0$, $\Delta_0 = \delta_0^T \delta_0$ is the well-known graph Laplacian
- $k = 1$, $\Delta_1 = \text{curl} \circ \text{curl}^* - \text{grad} \circ \text{div}$

Important Properties:
- $\Delta_k$ positive semi-definite
- $\ker(\Delta_k) = \ker(\delta_{k-1}^T) \cap \ker(\delta_k)$: divergence-free and curl-free, called harmonics
The space of pairwise rankings, $C^1(\chi_G, \mathbb{R})$, admits an orthogonal decomposition into three:

$$C^1(\chi_G, \mathbb{R}) = \text{im}(\delta_0) \oplus H_1 \oplus \text{im}(\delta^*_1)$$

where

$$H_1 = \ker(\delta_1) \cap \ker(\delta^*_0) = \ker(\Delta_1).$$
Figure: Hodge decomposition for pairwise rankings
Harmonic Rankings: Locally but NOT Globally Consistent

Figure: (a) a harmonic ranking; (b) from truncated Netflix network
Corollary

Every pairwise ranking admits a unique orthogonal decomposition,

\[ w = \text{proj}_{\text{im} \delta_0} w + \text{proj}_{\ker(\delta_0^*)} w \]

i.e.

pairwise = grad(global) + cyclic

Particularly the first projection grad(global) gives a global ranking

\[ x^* = (\delta_0^* \delta_0) \dagger \delta_0^* w = - (\Delta_0) \dagger \text{div} w \]
When Harmonic Ranking Vanishes

Corollary

If the clique complex $\chi_G$ has trivial 1-homology (no holes of boundary length $\geq 4$), then triangular transitivity ($\text{curl } w = 0$) implies the existence of global ranking $\text{im } \delta_0$.

- On simple-connected domains, curl-free vector fields are integrable.
- With trivial 1-homology $\chi_G$, local consistency implies the global consistency
- A particular case is when $G$ is a complete graph, the projection on transitivity subspace gives a unique global ranking, called Borda Count (1782) in social choice theory
Example: Erdős-Renyi Random Graph

**Theorem (Kahle ’06)**

For an Erdos-Renyi random graph $G(n, p)$ with $n$ vertices and each edge independently emerging with probability $p$, its clique complex $\chi_G$ almost always has zero 1-homology, except that

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

- Note that full Netflix movie-movie comparison graph is almost complete (0.22% missing edges), is that a Erdős-Renyi random graph?
Which Pairwise Ranking Model is More Consistent?

Curl distribution measures the **intrinsic local inconsistency** in a pairwise ranking:

![Curl Distribution of Pairwise Rankings](image)

**Figure:** Curl distribution of three pairwise rankings, based on most popular 500 movies. The pairwise score difference (the Linear model) in red have the thinnest tail.
Comparisons of Netflix Global Rankings

<table>
<thead>
<tr>
<th></th>
<th>Mean Score</th>
<th>Score Difference</th>
<th>Probability Difference</th>
<th>Logarithmic Odd Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Score</td>
<td>1.0000</td>
<td>0.9758</td>
<td>0.9731</td>
<td>0.9746</td>
</tr>
<tr>
<td>Score Difference</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9976</td>
<td>0.9977</td>
</tr>
<tr>
<td>Probability Difference</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9992</td>
<td>0.9992</td>
</tr>
<tr>
<td>Logarithmic Odd Ratio</td>
<td>-</td>
<td>6.03%</td>
<td>7.16%</td>
<td>7.15%</td>
</tr>
</tbody>
</table>

**Table:** Kendall’s rank correlation coefficients between different global rankings for Netflix. Note that the pairwise score difference (the Linear model) has the smallest relative residue.
Why Pairwise Ranking Works for Netflix?

- Pairwise rankings are good approximations of gradient flows on movie-movie networks.
- In fact, Netflix data in the large scale behave like a 1-dimensional curve in high dimensional space.
- To visualize this, we may use a spectral embedding approach.
Spectral Embedding

- Map every movie to a point in $S^4$ by
  \[
  \text{movie } m \rightarrow (\sqrt{p_1(m)}, \ldots, \sqrt{p_5(m)})
  \]
  where $p_k(m)$ is the probability that movie $m$ is rated as star $k \in \{1, \ldots, 5\}$. Obtain a movie-by-star matrix $Y$.

- Do SVD on $Y$, which is equivalent to do eigenvalue decomposition on the linear kernel
  \[
  K(s, t) = \langle s, t \rangle^d, \quad d = 1, \quad s, t \in S^4
  \]
  $K(s, t)$ is non-negative, whence the first eigenvector captures the centricity (density) of data and the second captures a tangent field of the manifold.
Figure: The second singular vector is monotonic to the mean score, indicating the intrinsic parameter of the horseshoe curve is driven by the mean score.
Conclusions

- Ranking as \textit{1-dimensional scaling} of data
- Pairwise ranking as approximate gradient fields or \textit{flows on graphs}
- Hodge Theory provides an \textit{orthogonal decomposition} for pairwise ranking flows
- This decomposition helps characterize the \textit{local (triangular)} vs. \textit{global consistency} of pairwise rankings, and gives a natural \textit{rank aggregation} scheme
- Geometry and topology play important roles, but everything is just \textit{linear algebra}!
Acknowledgements

- Gunnar Carlsson
- Vin de Silva
- Persi Diaconis
- Leo Guibas
- Fei Han
- Susan Holmes
- Qixing Huang
- Xiaoye Jiang
- Ming Ma
- Jason Morton
- Art Owen
- Michael Saunders
- Harlan Sexton
- Steve Smale
- Shmuel Weinberger
- Ya Xu