Applied Hodge Theory: Social Choice, Crowdsourced Ranking, and Game Theory

Yuan Yao

HKUST

Topological & Geometric Data Analysis

- Differential Geometric methods: manifolds
  - data manifold: manifold learning/NDR, etc.
  - model manifold: information geometry (high-order efficiency for parametric statistics), Grassmannian, etc.
- Algebraic Geometric methods: polynomials/varieties
  - data: tensor, Sum-Of-Square (MDS, polynom. optim.), etc
  - model: algebraic statistics
- Algebraic Topological methods: complexes (graphs, etc.)
  - persistent homology
  - *Euler calculus
  - Hodge theory (a bridge between geometry and topology via optimization/spectrum)

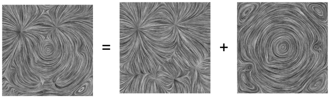
Helmholtz-Hodge Decomposition

Theorem (c.f. Marsden-Chorin 1992)

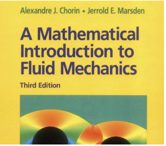
A vector field $\mathbf{w}$ *on a simply-connected* D *can be uniquely decomposed in the form*

$$\mathbf{w} = \mathbf{u} + \operatorname{grad} \phi$$

where $\mathbf{u}$ *has zero divergence and is parallel to* ∂D.



vector field = curl-free + div-free



Hodge Theory in Linear Algebra

For inner product spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, consider

$$\mathcal{X} \xrightarrow{A} \mathcal{Y} \xrightarrow{B} \mathcal{Z}.$$

and $\Delta = AA^* + B^*B : \mathcal{Y} \to \mathcal{Y}$ where $(\cdot)^*$ is adjoint operator of $(\cdot)$. If

$$B \circ A = 0,$$

then $\ker(\Delta) = \ker(A^*) \cap \ker(B)$ and *orthogonal* decomposition

$$\mathcal{Y} = \mathrm{im}(A) + \ker(\Delta) + \mathrm{im}(B^*)$$

Note: $\ker(B)/\,\mathrm{im}(A) \simeq \ker(\Delta)$ is the (real) (co)-homology group ($\mathbb{R} \to$ rings; vector spaces$\to$module).

Hodge Decomposition=Rank-Nullity Theorem

Proof in three lines:

$$
\begin{aligned}
\mathcal{Y} &= \operatorname{im}(A) + \ker(A^*), \qquad \text{by rank-nullity thm} \\
&= \operatorname{im}(A) + \ker(A^*)/\operatorname{im}(B^*) + \operatorname{im}(B^*), \\
&\qquad\qquad \text{since } \operatorname{im}(A) \subseteq \ker(B) \text{ by } BA = 0 \\
&\qquad\qquad \Leftrightarrow A^* B^* = 0 \Rightarrow \operatorname{im}(B^*) \subseteq \ker(A^*) \\
&= \operatorname{im}(A) + \ker(A^*) \cap \ker(B) + \operatorname{im}(B^*) \\
&\qquad\qquad \text{since } \ker(B) \oplus \operatorname{im}(B^*) = \mathcal{Y}
\end{aligned}
$$

Dirac Operator

Take product space $V = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, define

$$D = \begin{pmatrix} 0 & 0 & 0 \\ A & 0 & 0 \\ 0 & B & 0 \end{pmatrix}, \quad BA = 0,$$

Laplacian

$$L = (D+D^*)^2 = \mathrm{diag}(A^*A, AA^*+B^*B, BB^*) = \mathrm{diag}(L_0, L_1, L_2^{(down)})$$

Rank-nullity Theorem: $\mathrm{im}(D) + \ker(D^*) = V$, restricting on $\mathcal{Y}$ gives $\mathcal{Y} = \mathrm{im}(A) + \ker(A^*)$

Classical Hodge Theory on Riemannian Manifolds

- (W.V.D. Hodge, 1903-1975) de Rham complex:

$$0 \to \Omega^0(M) \xrightarrow{d_0} \Omega^1(M) \xrightarrow{d_1} \cdots \xrightarrow{d_{n-1}} \Omega^n(M) \xrightarrow{d_n} 0$$

  - M: compact Riemannian manifold
  - $\Omega^k(M)$: with k-differential forms
  - d: the exterior derivative operator

$$d^2 = d_k \circ d_{k-1} = 0$$

Hodge Theory on Metric Spaces

- (Alexander-Spanier, Bartholdi-Schick-Smale-Smale, 2011)

$$0 \to L^2(X) \xrightarrow{d_0} L^2(X^2) \xrightarrow{d_1} \cdots \xrightarrow{d_{n-1}} L^2(X^n) \xrightarrow{d_n} \cdot$$

  - X: metric space
  - $L^2(X)$: square integral functions on X
  - $d : L^2(X^k) \to L^2(X^{k+1})$ – finite difference

$$(df)(x_0, \ldots, x_k) = \sum_{i=1}^{k} (-1)^i \prod_{j \neq i} \sqrt{K(x_i, x_j)} f(x_{-i})$$

  - adjoint operator $\delta : L^2(X^{k+1}) \to L^2(X^k)$

$$\delta g(x) = \sum_{i=0}^{k} (-1)^i \int_X \prod_{j=0}^{k-1} \sqrt{K(t, x_j)} g(x_0, \ldots, x_{i-1}, t, x_i, \ldots, x_{k-1}) dt$$

Combinatorial Hodge Theory on Simplicial Complexes

$$0 \to \Omega^0(X) \xrightarrow{d_0} \Omega^1(X) \xrightarrow{d_1} \cdots \xrightarrow{d_{n-1}} \Omega^n(X) \xrightarrow{d_n} \cdots$$

- X is finite
- $\chi(X) \subseteq 2^X$: simplicial complex formed by X $\Leftrightarrow$ if $\tau \in \chi(X)$ and $\sigma \subseteq \tau$, then $\sigma \in \chi(X)$
- k-forms or cochains as alternating functions

$$\Omega^k(X) = \{u : \chi_{k+1}(X) \to \mathbb{R}, u_{i_{\sigma(0)},\ldots,i_{\sigma(k)}} = \text{sign}(\sigma) u_{i_0,\ldots,i_k}\}$$

- coboundary maps $d_k : \Omega^k(X) \to \Omega^{k+1}(X)$ alternating difference

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

- $d_k \circ d_{k-1} = 0$

Example: graph and clique complex

- $G = (X, E)$ is a undirected but oriented graph
- Clique complex $\chi_G \subseteq 2^X$ collects all complete subgraph of G
- k-forms or cochains $\Omega^k(\chi_G)$ as alternating functions:
  - 0-forms: $v : V \to \mathbb{R} \cong \mathbb{R}^n$
  - 1-forms as skew-symmetric functions: $w_{ij} = -w_{ji}$
  - 2-forms as triangular-curl:

  $z_{ijk} = z_{jki} = z_{kij} = -z_{jik} = -z_{ikj} = -z_{kji}$
- coboundary operators as alternating difference operators:
  - $(d_0 v)(i,j) = v_j - v_i =: (\mathrm{grad}\, v)(i,j)$
  - $(d_1 w)(i,j,k) = (\pm)(w_{ij} + w_{jk} + w_{ki}) =: (\mathrm{curl}\, w)(i,j,k)$
- $d_1 \circ d_0 = \mathrm{curl}(\mathrm{grad}\, u) = 0$

Hodge Laplacian

- combinatorial Laplacian $\Delta = d_{k-1} d_{k-1}^* + d_k^* d_k$
  - $k = 0$, $\Delta_0 = d_0^* d_0$ is the (unnormalized) graph Laplacian
  - $k = 1$, 1-Hodge Laplacian (Helmholtzian)

$$\Delta_1 = \mathrm{curl} \circ \mathrm{curl}^* - \mathrm{div} \circ \mathrm{grad}$$

- Hodge decomposition holds for $\Omega^k(X)$
  - $\Omega^k(X) = \mathrm{im}(d_{k-1}) \oplus \ker(\Delta_k) \oplus \mathrm{im}(\delta_k)$
  - $\dim(\Delta_k) = \beta_k(\chi(X))$

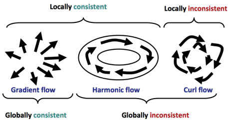


Figure: Courtesy by Asu Ozdaglar

Particularly when $k = 1$

Discrete de Rham complex:

$$\Omega^0 \xrightarrow{d_0} \Omega^1 \xrightarrow{d_1} \Omega^2,$$

i.e.

$$\text{Potential} \xrightarrow{\text{grad}} \text{Edge-flow} \xrightarrow{\text{curl}} \text{Triangular-curl}$$

$$\text{Potential} \xleftarrow{\text{grad}^*(=:-\text{div})} \text{Edge-flow} \xleftarrow{\text{curl}^*} \text{Triangular-curl}.$$

Note that

$$\text{curl} \circ \text{grad}(\text{Potential}) = 0 \Leftrightarrow d_1 d_0 = 0.$$

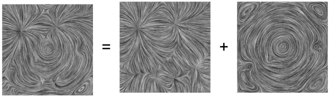
Hodge decomposition

$$\text{Edge-flow} = \text{grad}(\text{Potential}) \oplus \text{harmonic} \oplus \text{curl}^*(\text{Triangular})$$

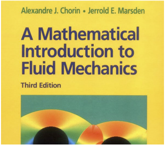
Helmholtz-Hodge Decomposition

When the first Betti number $\beta_1(\chi_G) = 0$, *harmonic* vanishes and (combinatorial/discrete) Helmholtz-Hodge decomposition:

$$\text{Edge-flow} = \text{grad}(\text{Potential}) \oplus \text{curl}^*(\text{Triangular})$$



vector field = curl-free + div-free



Combinatorial Hodge theory: matrix version

A skew-symmetric matrix W associated with G (with missing values) can be decomposed uniquely

$$W = W_1 + W_2 + W_3$$

where

- W_1 satisfies
  - '*integrable*': $W_1(i,j) = v_j - v_i$ for some $v : V \to \mathbb{R}$.
- W_2 satisfies
  - '*curl free*': $W_2(i,j) + W_2(j,k) + W_2(k,i) = 0$ for all 3-clique (i,j,k);
  - '*divergence free*': $\sum_{j:(i,j)\in E} W_2(i,j) = 0$
- $W_3 \perp W_1$ and $W_3 \perp W_2$, which is divergence-free but not curl-free

Forgetful functors

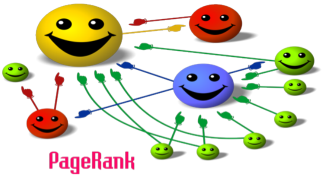
Riemannian manifolds $\rightarrow$ Metric spaces $\rightarrow$ Cell complexes

- From differentiable to combinatorial structures, Hodge decomposition is universal
- Topological invariants (homology) are preserved in such coarse-grained functors
- Natural for data analysis, a connection between geometry and topology: harmonic basis

Social Choice Problem

The fundamental problem of preference aggregation:

How to aggregate preferences
which faithfully represent individuals?



Thurstone's Crime Scaling in 1928

TABLE 1.

TABLE OF PROPORTIONS p_{ka} OR $p_{k>a}$

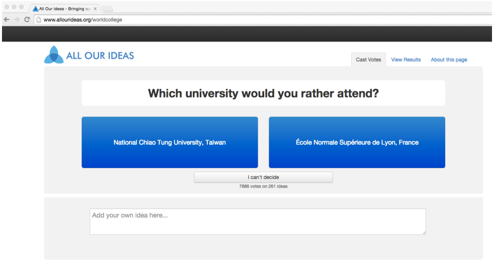
	Abortion	Adultery	Arson	Assault and battery	Bootlegging	Burglary	Counterfeiting	Embezzlement	Forgery	Homicide	Kidnapping	Larceny	Libel	Perjury	Rape	Receiving stolen goods	Seduction	Smuggling	Vagrancy
Abortion	....	.323	.338	.211	.128	.238	.244	.245	.212	.760	.318	.222	.191	.256	.822	.143	.419	.174	.045
Adultery	.677	....	.415	.242	.172	.281	.285	.253	.274	.863	.365	.207	.182	.245	.925	.143	.589	.204	.034
Arson	.662	.585	....	.260	.136	.226	.321	.348	.254	.017	.563	.215	.144	.349	.944	.140	.716	.170	.019
Assault and battery	.789	.757	.740	....	.379	.515	.556	.485	.534	.070	.743	.385	.385	.587	.947	.344	.785	.346	.072
Bootlegging	.872	.828	.864	.621	....	.764	.745	.738	.754	.955	.924	.678	.506	.728	.985	.527	.871	.576	.116
Burglary	.762	.719	.774	.485	.236	....	.593	.605	.580	.981	.856	.333	.322	.478	.981	.221	.769	.284	.027
Counterfeiting	.756	.715	.679	.444	.255	.407	....	.540	.488	.947	.804	.303	.284	.532	.963	.199	.756	.215	.042
Embezzlement	.755	.747	.652	.515	.262	.395	.460	....	.350	.958	.752	.305	.248	.474	.977	.141	.774	.251	.049
Forgery	.788	.726	.746	.466	.246	.420	.512	.650	....	.951	.819	.343	.320	.534	.966	.195	.820	.260	.035
Homicide	.240	.137	.083	.030	.045	.019	.053	.042	.049	....	.083	.030	.034	.079	.441	.027	.181	.026	.011
Kidnapping	.682	.635	.437	.257	.076	.144	.196	.248	.181	.917	....	.170	.106	.288	.902	.098	.595	.086	.026
Larceny	.778	.793	.785	.615	.322	.667	.697	.695	.657	.970	.830	....	.348	.648	.970	.268	.848	.365	.053
Libel	.809	.818	.855	.615	.494	.678	.716	.752	.680	.966	.894	.652	....	.702	.981	.530	.886	.456	.067
Perjury	.744	.755	.651	.413	.272	.522	.467	.526	.466	.921	.712	.352	.298	....	.951	.204	.767	.222	.015
Rape	.178	.075	.056	.053	.015	.019	.037	.023	.034	.559	.098	.030	.019	.049	....	.019	.076	.023	.015
Receiv'g stolen goods	.857	.857	.860	.656	.473	.779	.801	.859	.805	.973	.902	.732	.470	.796	.981	....	.875	.525	.061
Seduction	.581	.411	.284	.215	.129	.231	.244	.226	.180	.819	.405	.152	.114	.233	.924	.125	....	.121	.023
Smuggling	.826	.796	.830	.654	.424	.716	.785	.749	.740	.974	.914	.635	.544	.778	.977	.475	.879	....	.037
Vagrancy	.955	.966	.981	.928	.884	.973	.958	.951	.965	.989	.974	.947	.933	.985	.985	.939	.977	.963	....

Figure: Can we learn a scale for crimes (wine taste) from paired comparison?

Crowdsourcing QoE evaluation of Multimedia

Figure: Crowdsouring subjective Quality of Experience evaluation (Xu-Huang-Y., et al. *ACM-MM* 2011)

Crowdsourced ranking



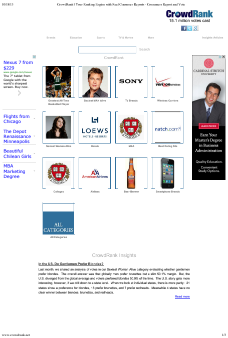


Figure: Left: www.allourideas.org/worldcollege (Prof. Matt Salganik at Princeton); Right: www.crowdrank.net.

Learning relative attributes: age

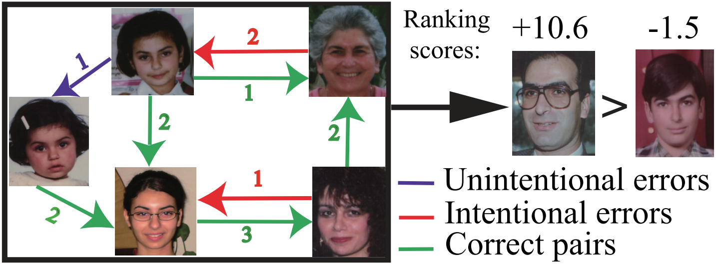


Figure: Age: a relative attribute estimated from paired comparisons (Fu-Hospedales-Xiang-Gong-Y. *ECCV*, 2014)

Language Models with Human Preference Feedback

- Large language models need alignment with human preference
- Human preference is noisy, relatively reliable in pairwise comparison
- Reinforcement Learning with a reward model from human preference feedback (RLHF)

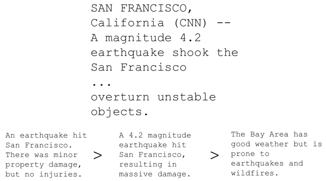


Figure: News and Summaries from Language Models, with human preference order

Netflix Customer-Product Rating

Example (Netflix Customer-Product Rating)

- 480189-by-17770 customer-product 5-star rating matrix X with $X_{ij} = \{1, \ldots, 5\}$
- X contains 98.82% missing values

However,

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

Drug Sensitivity Ranking

Example (Drug Sensitivity Data)

- 300 drugs
- 940 cell lines, with ≈ 1000 genetic features
- sensitivity measurements in terms of IC50 and AUC
- heterogeneous missing values

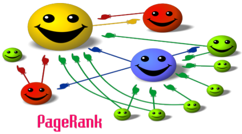
However,

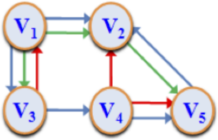
- every two drug d_1 and d_2 has been tested at least in one cell line, hence comparable (which is more sensitive)
- complete graph of paired comparisons: $G = (V, E)$
- imbalanced: number of raters on $e \in E$ varies

Paired comparison data on graphs

Graph $G = (V, E)$

- V: alternatives to be ranked or rated
- $(i_\alpha, j_\alpha) \in E$ a pair of alternatives
- $y_{ij}^\alpha \in \mathbb{R}$ degree of preference by rater α
- $\omega_{ij}^\alpha \in \mathbb{R}_+$ confidence weight of rater α
- Examples: relative attributes, subjective QoE assessment, perception of illuminance intensity, sports, wine taste, etc.





Modern settings

Modern ranking data are

- distributive on networks
- incomplete with missing values
- imbalanced
- even adaptive to dynamic and random settings?

Here we introduce:

Hodge Theory approach to Social Choice or Preference Aggregation

History

Classical social choice theory origins from Voting Theory

- *Borda* 1770, B. Count against plurality vote
- *Condorcet* 1785, C. Winner who wins all paired elections
- Impossibility theorems: *Kenneth Arrow* 1963, *Amartya Sen* 1973
- Resolving conflicts: *Kemeny*, *Saari* ...
- In these settings, we study complete ranking orders from voters.

Classical Social Choice or Voting Theory

Problem

Given m voters whose preferences are total orders (permutation) $\{\succeq_i : i = 1, \ldots, m\}$ *on a candidate set* V*, find a social choice mapping*

$$f : (\succeq_1, \ldots, \succeq_m) \mapsto \succeq^*,$$

as a total order on V*, which "best" represents voter's will.*

Example: 3 candidates ABC

Preference order	Votes
$A \succeq B \succeq C$	2
$B \succeq A \succeq C$	3
$B \succeq C \succeq A$	1
$C \succeq B \succeq A$	3
$C \succeq A \succeq B$	2
$A \succeq C \succeq B$	2

What we did in practice I: Position rules

There are two important classes of social mapping in realities:

- I. Position rules: assign a score $s : V \to \mathbb{R}$, such that for each voter's order(permutation) $\sigma_i \in S_n$ $(i = 1, \ldots, m)$, $s_{\sigma_i(k)} \geq s_{\sigma_i(k+1)}$. Define the social order by the descending order of total score over raters, i.e. the score for k-th candidate

$$f(k) = \sum_{i=1}^{m} s_{\sigma_i}(k).$$

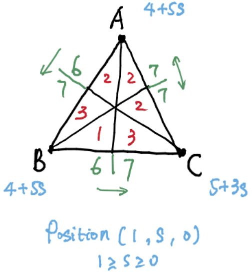
  - Borda Count: $s : V \to \mathbb{R}$ is given by $(n-1, n-2, \ldots, 1, 0)$
  - Vote-for-top-1: $(1, 0, \ldots, 0)$
  - Vote-for-top-2: $(1, 1, 0, \ldots, 0)$

What we did in practice II: pairwise rules

- II. Pairwise rules: convert the voting profile, a (distribution) function on $n!$ set S_n, into paired comparison matrix $X \in \mathbb{R}^{n \times n}$ where $X(i,j)$ is the number (distribution) of voters that $i \succ j$; define the social order based on paired comparison data X.
  - Kemeny Optimization: minimizes the number of pairwise mismatches to X over S_n (NP-hard)
  - Pluarity: the number of wins in paired comparisons (tournaments) – equivalent to Borda count in complete Round-Robin tournaments

Revisit the ABC-Example

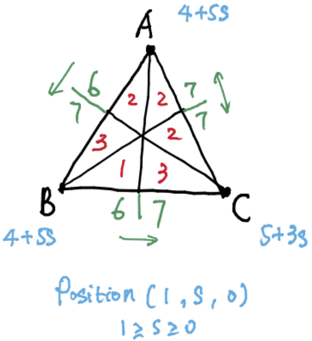
Preference order	Votes
$A \succeq B \succeq C$	2
$B \succeq A \succeq C$	3
$B \succeq C \succeq A$	1
$C \succeq B \succeq A$	3
$C \succeq A \succeq B$	2
$A \succeq C \succeq B$	2



Voting chaos!

- Position:
  - $s < 1/2$, C wins
  - $s = 1/2$, ties
  - $s > 1/2$, A, B wins
- Pairwise:
  - A, B: 13 wins
  - C: 14 wins
  - Condorcet winner: C

so completely in chaos!



Arrow's Impossibility Theorem

(Arrow'1963)

Consider the Unrestricted Domain, i.e. voters may have all complete and transitive preferences. The only social choice rule satisfying the following conditions is the dictator rule

- Pareto (Unanimity): if all voters agree that $A \succeq B$ then such a preference should appear in the social order
- Independence of Irrelevant Alternative (IIA): the social order of any pair only depends on voter's relative rankings of that pair

Sen's Impossibility Theorem

(Sen'1970)

With Unrestricted Domain, there are cases with voting data that no social choice mapping,

$$f : (\succeq_1, \ldots, \succeq_m) \mapsto 2^V,$$

exists under the following conditions

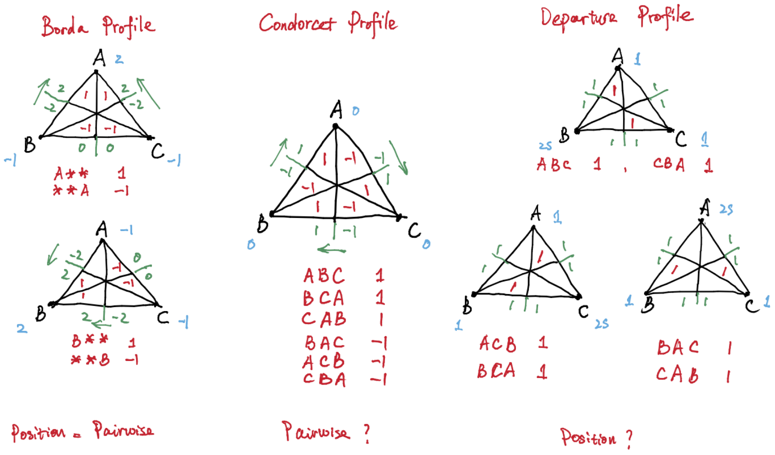
- Pareto: if all voters agree that $A > B$ then such a preference should appear in the social order
- Minimal Liberalism: two distinct voters decide social orders of two distinct pairs respectively

A Possibility: Saari's Profile Decomposition

Every voting profile, as distributions on symmetric group S_n, can be decomposed into the following components:

- Universal kernel: all ranking methods induce a complete tie on any subset of V
  - dimension: $n! - 2^{n-1}(n-2) - 2$
- Borda profile: all ranking methods give the same result
  - dimension: $n - 1$
  - basis: $\{1(\sigma(1) = i, *) - 1(*, \sigma(n) = i) : i = 1, \ldots, n\}$
- Condorcet profile: all positional rules give the same result
  - dimension: $\frac{(n-1)!}{2}$
  - basis: sum of Z_n orbit of σ minus their reversals
- Departure profile: all pairwise rules give the same result

Example: Decomposition of Voting Profile $R^{3!}$



Borda Count: the most consistent rule?

Table: Invariant subspaces of social rules (-)

	Borda Profile	Condorcet	Departure
Borda Count	consistent	-	-
Pairwise	consistent	inconsistent	-
Position (non-Borda)	consistent	-	inconsistent

- So, if you look for a best possibility from impossibility, Borda count is perhaps the choice
- Borda Count is the projection onto the Borda Profile subspace

Equivalently, Borda Count is a Least Square

Borda Count is equivalent to

$$\min_{\beta \in \mathbb{R}^{|V|}} \sum_{\alpha, \{i,j\} \in E} \omega_{ij}^{\alpha} (\beta_i - \beta_j - Y_{ij}^{\alpha})^2,$$

where

- E.g. $Y_{ij}^{\alpha} = 1$, if $i \succeq j$ by voter α, and $Y_{ij}^{\alpha} = -1$, on the opposite.
- Note: NP-hard ($n > 3$) Kemeny Optimization, or Minimimum-Feedback-Arc-Set:

$$\min_{s \in \mathbb{R}^{|V|}} \sum_{\alpha, \{i,j\} \in E} \omega_{ij}^{\alpha} (\mathrm{sign}(\beta_i - \beta_j) - \hat{Y}_{ij}^{\alpha})^2$$

Generalized Borda Count with Incomplete Data

$$\min_{x\in \mathbb{R}^{|V|}} \sum_{\alpha,\{i,j\}\in E} \omega_{ij}^{\alpha}(x_i - x_j - y_{ij}^{\alpha})^2,$$

$$\Leftrightarrow$$

$$\min_{x\in \mathbb{R}^{|V|}} \sum_{\{i,j\}\in E} \omega_{ij}((x_i - x_j) - \hat{y}_{ij})^2,$$

$$\text{where } \hat{y}_{ij} = \hat{\mathbb{E}}_{\alpha} y_{ij}^{\alpha} = (\sum_{\alpha} \omega_{ij}^{\alpha} y_{ij}^{\alpha})/\omega_{ij} = -\hat{y}_{ji}, \quad \omega_{ij} = \sum_{\alpha} \omega_{ij}^{\alpha}$$

So $\hat{y} \in l^2_{\omega}(E)$, inner product space with $\langle u, v\rangle_{\omega} = \sum u_{ij} v_{ij} \omega_{ij}$, u, v skew-symmetric

Statistical Majority Voting: $l^2(E)$

- $\hat{y}_{ij} = (\sum_\alpha \omega_{ij}^\alpha y_{ij}^\alpha)/(\sum_\alpha \omega_{ij}^\alpha) = -\hat{y}_{ji}$, $\omega_{ij} = \sum_\alpha \omega_{ij}^\alpha$
- $\hat{y}$ from generalized linear models:
  - [1] *Uniform* model: $\hat{y}_{ij} = 2\hat{\pi}_{ij} - 1$.
  - [2] *Bradley-Terry* model: $\hat{y}_{ij} = \log \frac{\hat{\pi}_{ij}}{1-\hat{\pi}_{ij}}$.
  - [3] *Thurstone-Mosteller* model: $\hat{y}_{ij} = \Phi^{-1}(\hat{\pi}_{ij})$, $\Phi(x)$ is Gaussian CDF
  - [4] *Angular transform* model: $\hat{y}_{ij} = \arcsin(2\hat{\pi}_{ij} - 1)$.

Hodge Decomposition of Pairwise Ranking

$\hat{y}_{ij} = -\hat{y}_{ji} \in l^2_\omega(E)$ admits an orthogonal decomposition,

$$\hat{y} = Ax + B^T z + w, \quad (1)$$

where

$$(Ax)(i,j) := x_i - x_j, \quad \text{gradient, as Borda profile,} \quad (2a)$$

$$(B\hat{y})(i,j,k) := \hat{y}_{ij} + \hat{y}_{jk} + \hat{y}_{ki}, \quad \text{trianglar cycle/curl, Condorcet} \quad (2b)$$

$$w \in \ker(A^T) \cap \ker(B), \quad \text{harmonic, Condorcet.} \quad (2c)$$

In other words

$$\operatorname{im}(A) \oplus \ker(AA^T + B^T B) \oplus \operatorname{im}(B^T)$$

Why? Hodge Decomposition in Linear Algebra

For inner product spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, consider

$$\mathcal{X} \xrightarrow{A} \mathcal{Y} \xrightarrow{B} \mathcal{Z}.$$

and $\Delta = AA^* + B^*B : \mathcal{Y} \to \mathcal{Y}$ where $(\cdot)^*$ is adjoint operator of $(\cdot)$. If

$$B \circ A = 0,$$

then $\ker(\Delta) = \ker(A^*) \cap \ker(B)$ and *orthogonal* decomposition

$$\mathcal{Y} = \text{im}(A) + \ker(\Delta) + \text{im}(B^*)$$

Note: $\ker(B)/\,\text{im}(A) \simeq \ker(\Delta)$ is the (real) (co)-homology group ($\mathbb{R} \to$ rings; vector spaces$\to$module).

Hodge Decomposition=Rank-Nullity Theorem

Proof in three lines:

$$\begin{aligned}
\mathcal{Y} &= \operatorname{im}(A) + \ker(A^*), \qquad \text{by rank-nullity thm} \\
&= \operatorname{im}(A) + \ker(A^*)/\operatorname{im}(B^*) + \operatorname{im}(B^*), \\
&\qquad\qquad \text{since } \operatorname{im}(A) \subseteq \ker(B) \text{ by } BA = 0 \\
&= \operatorname{im}(A) + \ker(A^*) \cap \ker(B) + \operatorname{im}(B^*)
\end{aligned}$$

Dirac Operators and Laplacians

Take product space $V = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, define

$$D = \begin{pmatrix} 0 & 0 & 0 \\ A & 0 & 0 \\ 0 & B & 0 \end{pmatrix}, \quad BA = 0,$$

Laplacian

$$L = (D+D^*)^2 = \mathrm{diag}(A^*A, AA^*+B^*B, BB^*) = \mathrm{diag}(\Delta_0, \Delta_1, \Delta_2^{(down)})$$

where

- Δ_0 is the (unnormalized Graph Laplacian)
- Δ_1 is the Hodge 1-Laplacian
- $\Delta_2 : \mathcal{Z} \to \mathcal{Z}$ is the (downward) Hodge 2-Laplacian

Hence, in our case

Note $B \circ A = 0$ since

$$(B \circ Ax)(i,j,k) = (x_i - x_j) + (x_j - x_k) + (x_k - x_i) = 0.$$

Hence

$$A^T \hat{y} = A^T(Ax + B^T z + w) = A^T Ax \Rightarrow x = (\Delta_0)^\dagger A^T \hat{y}$$

$$B\hat{y} = B(Ax + B^T z + w) = BB^T z \Rightarrow z = (\Delta_2)^\dagger B\hat{y}$$

$$A^T w = Bw = 0 \Rightarrow w \in \ker(\Delta_1), \quad \Delta_1 = AA^T + B^T B.$$

Exterior Calculus on Simplicial Complexes

$$0 \to \Omega^0(X) \xrightarrow{d_0} \Omega^1(X) \xrightarrow{d_1} \cdots \xrightarrow{d_{n-1}} \Omega^n(X) \xrightarrow{d_n} \cdots$$

- X is finite
- $\chi(X) \subseteq 2^X$: simplicial complex formed by X $\Leftrightarrow$ if $\tau \in \chi(X)$ and $\sigma \subseteq \tau$, then $\sigma \in \chi(X)$
- k-forms or cochains as alternating functions

$$\Omega^k(X) = \{u : \chi_{k+1}(X) \to \mathbb{R}, u_{i_{\sigma(0)},\ldots,i_{\sigma(k)}} = \mathrm{sign}(\sigma) u_{i_0,\ldots,i_k}\}$$

- coboundary maps $d_k : \Omega^k(X) \to \Omega^{k+1}(X)$ alternating difference

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

- $d_k \circ d_{k-1} = 0$

Example: graph and clique complex

- $G = (X, E)$ is a undirected but oriented graph
- Clique complex $\chi_G \subseteq 2^X$ collects all complete subgraph of G
- k-forms or cochains $\Omega^k(\chi_G)$ as alternating functions:
  - 0-forms: $v : V \to \mathbb{R} \cong \mathbb{R}^n$
  - 1-forms as skew-symmetric functions: $w_{ij} = -w_{ji}$
  - 2-forms as triangular-curl:

  $$z_{ijk} = z_{jki} = z_{kij} = -z_{jik} = -z_{ikj} = -z_{kji}$$
- coboundary operators as alternating difference operators:
  - $(d_0 v)(i,j) = v_j - v_i =: (\mathrm{grad}\, v)(i,j)$
  - $(d_1 w)(i,j,k) = (\pm)(w_{ij} + w_{jk} + w_{ki}) =: (\mathrm{curl}\, w)(i,j,k)$
- $d_1 \circ d_0 = \mathrm{curl}(\mathrm{grad}\, u) = 0$

Hodge Laplacian and Decomposition

- combinatorial Laplacian $\Delta = d_{k-1}d_{k-1}^* + d_k^* d_k$
  - $k = 0$, $\Delta_0 = d_0^* d_0$ is the (unnormalized) graph Laplacian
  - $k = 1$, 1-Hodge Laplacian (Helmholtzian)

$$\Delta_1 = \mathrm{curl} \circ \mathrm{curl}^* - \mathrm{div} \circ \mathrm{grad}$$

- Hodge decomposition holds for $\Omega^k(X)$
  - $\Omega^k(X) = \mathrm{im}(d_{k-1}) \oplus \ker(\Delta_k) \oplus \mathrm{im}(\delta_k)$
  - $\dim(\ker(\Delta_k)) = \beta_k(\chi(X))$, k-harmonics

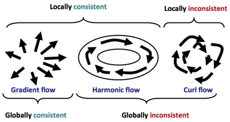


Figure: Courtesy by Asu Ozdaglar

Generalized Borda Count estimator

Gradient flow $\hat{y}^{(g)} := (Ax)(i,j) = x_i - x_j$ gives the generalized Borda count score, x which solves Graph Laplacian equation

$$\min_{x \in \mathbb{R}^{|V|}} \sum_{\alpha,(i,j)\in E} \omega_{ij}^{\alpha}(x_i - x_j - y_{ij}^{\alpha})^2 \Leftrightarrow \Delta_0 x = A^T \hat{y}$$

where $\Delta_0 = A^T A$ is the unnormalized graph Laplacian of G.

- In theory, nearly linear algorithms for such equations, e.g. Spielman-Teng'04, Koutis-Miller-Peng'12, etc.
- But in practice? ...

Online HodgeRank [Xu-Huang-Yao'2012]

Robbins-Monro (1951) algorithm for $\Delta_0 x = \bar{b} := \delta_0^* \hat{y}$,

$$x_{t+1} = x_t - \gamma_t (A_t x_t - b_t), \quad x_0 = 0,\ \mathbb{E}(A_t) = \Delta_0,\ \mathbb{E}(b_t) = \bar{b}$$

Note:

- For each $Y_t(i_{t+1}, j_{t+1})$, updates only occur locally
- Step size: $\gamma_t = a(t+b)^{-1/2}$ (e.g. a=$1/\lambda_1(\Delta_0)$ and b large)
- Optimal convergence of x_t to x^* (population solution) in t

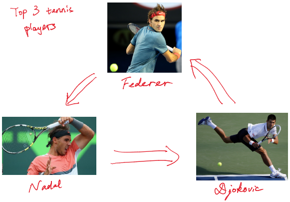
$$\mathbb{E}\|x_t - x^*\|^2 \leq O\left(t^{-1} \cdot \lambda_2^{-2}(\Delta_0)\right)$$

  where $\lambda_2(\Delta_0)$ is the Fiedler Value of graph Laplacian
- Tong Zhang's SVRG: $\mathbb{E}\|s_t - s^*\|^2 \leq O\left(t^{-1} + \lambda_2^{-2}(\Delta_0) t^{-2}\right)$

Condorcet Profile splits into Local vs. Global Cycles

Residues $\hat{y}^{(c)} = B^T z$ and $\hat{y}^{(h)} = w$ are cyclic rankings, accounting for conflicts of interests:

- $\hat{y}^{(c)}$, the local/triangular inconsistency, triangular curls (Z_3-invariant)
  - $\hat{y}^{(c)}_{ij} + \hat{y}^{(c)}_{jk} + \hat{y}^{(c)}_{ki} \neq 0$, $\{i, j, k\} \in T$



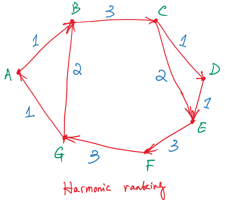
Condorcet Profile in Harmonic Ranking

- $\hat{y}^{(h)} = w$, the global inconsistency, harmonic ranking (Z_n-invariant)

$$\hat{y}^{(h)}_{ij} + \hat{y}^{(h)}_{jk} + \hat{y}^{(h)}_{ki} = 0, \quad \text{for each } \{i,j,k\} \in T, \tag{3a}$$

$$\sum_{j\sim i} \omega_{ij}\hat{y}^{(h)}_{ij} = 0, \quad \text{for each } i \in V. \tag{3b}$$

- voting chaos: *circular coordinates* on $V \Rightarrow$ *fixed tournament* issue



Cyclic Ranking and Outliers: High Dimensional Statistics

- Outliers are sparse approximation of cyclic rankings (curl+harmonic) [Xu-Xiong-Huang-Y.'13]

$$\min_{\gamma} \|\Pi_{\ker(A^*)}(\hat{y} - \gamma)\|^2 + \lambda \|\gamma\|_1$$

- Robust ranking can be formulated as a Huber's LASSO

$$\min_{x,\gamma} \|\hat{y} - Ax - \gamma\|^2 + \lambda \|\gamma\|_1$$

  - outlier γ is incidental parameter (Neyman-Scott'1948)
  - global rating x is structural parameter
- Yet, LASSO is a biased estimator (Fan-Li'2001)

A Differential Inclusion Approach to Sparse Learning

- A Dual Gradient Descent (sparse mirror descent) dynamics [Osher-Ruan-Xiong-Y.-Yin'2014, Huang-Sun-Xiong-Y.'2020]

$$\dot{\rho}_t = \frac{1}{n} X^T (y - X\beta_t), \tag{4a}$$

$$\rho_t \in \partial \|\beta_t\|_1. \tag{4b}$$

- called Inverse Scale Space dynamics in imaging
- sign consistency under nearly the same conditions as LASSO (Wainwright'99), yet returns unbiased estimator
- fast and scalable discretization as linearized Bregman Iteration

Conflicts are due to personalization [Xu-...-Y.'2019]

$$\textit{cycles} = \textit{personalized ranking} + \textit{position bias} + \textit{noise}.$$

Linear mixed-effects model for annotator's pairwise ranking:

$$y_{ij}^u = (\theta_i + \delta_i^u) - (\theta_j + \delta_j^u) + \gamma^u + \varepsilon_{ij}^u, \qquad (5)$$

where

- θ_i is the common global ranking score, as a fixed effect;
- δ_i^u is the annotator's preference deviation from the common ranking θ_i such that $\theta_i^u := \theta_i + \delta_i^u$ is u's personalized ranking;
- γ^u is an annotator's position bias, which captures the careless behavior by clicking one side during the comparisons;
- ε_{ij}^u is the random noise which is assumed to be independent and identically distributed with zero mean and being bounded.

Movielens Multilevel Rankings

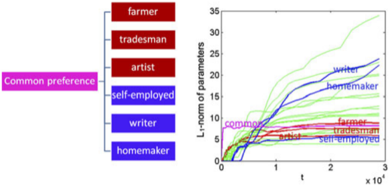


Figure: A two-level preference learning in MovieLens: (a) The common preference with six representative occupation group preference. (b) The purple is the common preference, the remaining 21 paths represent the occupation group preferences, the red are the three groups with most distinct preferences from the common, the blue are the three groups with most similar preferences to the common, and the green ones are the others [Xu-Xiong-Huang-Cao-Y.'2019].

Topological Obstructions

Two topological conditions are important:

- Connectivity:
  - G is connected $\Rightarrow$ unique global ranking is possible;
- Loop-free:
  - for cyclic rankings, consider clique complex $\chi_G^2 = (V, E, T)$ by attaching triangles $T = \{(i, j, k)\}$
  - $\dim(\ker(\Delta_1)) = \beta_1(\chi_G^2)$, so harmonic ranking $w = 0$ if χ_G^2 is loop-free, here topology plays a role of obstruction of fixed-tournament
  - "Triangular arbitrage-free implies arbitrage-free"

Persistent Homology: online algorithm for topology tracking (e.g Edelsbrunner-Harer'08)

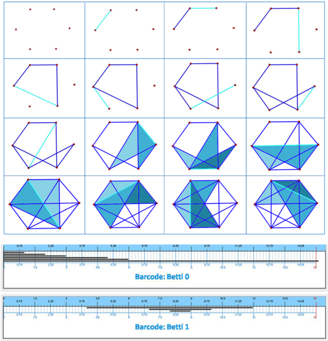


Figure: Persistent Homology Barcodes

- vertice, edges, and triangles etc. sequentially added
- online update of homology
- $O(m)$ for surface embeddable complex; and $O(m^{2.xx})$ in general (m number of simplex)

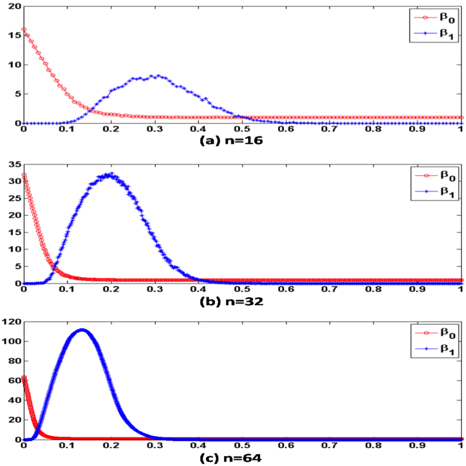
Random Graph Models for Crowdsourcing

- Recall that in crowdsourcing ranking on internet,
  - unspecified raters compare item pairs randomly
  - online, or sequentially sampling
- random graph models for experimental designs
  - P a distribution on random graphs, invariant under permutations (relabeling)
  - Generalized de Finetti's Theorem [Aldous 1983, Kallenberg 2005]: $P(i,j)$ (P ergodic) is an uniform mixture of

$$h(u,v) = h(v,u) : [0,1]^2 \to [0,1],$$

  h unique up to sets of zero-measure
  - Erdös-Rényi: $P(i,j) = P(edge) = \int_0^1 \int_0^1 h(u,v)dudv =: p$
  - edge-independent process (Chung-Lu'06)

Phase Transitions in Erdös-Rényi Random Graphs



Phase Transitions of Large Random Graphs

For an Erdos-Renyi random graph $G(n, p)$ with n vertices and each edge independently emerging with probability $p(n)$,

- (Erdös-Rényi 1959) One phase-transition for β_0
  - $p << 1/n^{1+\epsilon}$ $(\forall\epsilon > 0)$, almost always disconnected
  - $p >> log(n)/n$, almost always connected
- (Kahle 2009) Two phase-transitions for β_k $(k \geq 1)$
  - $p << n^{-1/k}$ or $p >> n^{-1/(k+1)}$, almost always β_k vanishes;
  - $n^{-1/k} << p << n^{-1/(k+1)}$, almost always β_k is nontrivial

For example: with $n = 16$, 75% distinct edges included in G, then χ_G with high probability is connected and loop-free. In general, $O(n\log(n))$ samples for connectivity and $O(n^{3/2})$ for loop-free.

Three sampling methods

- *Uniform sampling with replacement (i.i.d.)* $(G_0(n,m))$.
  - Each edge is sampled from the uniform distribution on $\binom{n}{2}$ edges, with replacement. This is a weighted graph and the sum of weights is m.
- *Uniform sampling without replacement* $(G(n,m))$.
  - Each edge is sampled from the uniform distribution on the available edges without replacement. For $m \leq \binom{n}{2}$, this is an instance of the Erdös-Rényi random graph model $G(n,p)$ with $p = m/\binom{n}{2}$.
- *Greedy sampling* $(G_\star(n,m))$.
  - Each pair is sampled to maximize the algebraic connectivity of the graph in a greedy way: the graph is built iteratively; at each iteration, the Fiedler vector is computed and the edge (i,j) which maximizes $(\psi_i - \psi_j)^2$ is added to the graph.

Key Estimates of Fiedler Value near Connectivity Threshold.

$$G_0(n,m):\ \ \frac{\lambda_2}{np} \approx a_1(p_0,n) := 1 - \sqrt{\frac{2}{p_0}}\sqrt{1-\frac{2}{n}} \qquad (6)$$

$$G(n,m):\ \ \frac{\lambda_2}{np} \approx a_2(p_0,n) := 1 - \sqrt{\frac{2}{p_0}}\sqrt{1-p} \qquad (7)$$

where $p_0 := 2m/(n\log n) \geq 1$, $p = \frac{p_0 \log n}{n}$ and

$$a(p_0) = 1 - \sqrt{2/p_0} + O(1/p_0), \quad \text{for } p_0 \gg 1.$$

Without-replacement as good as Greedy!

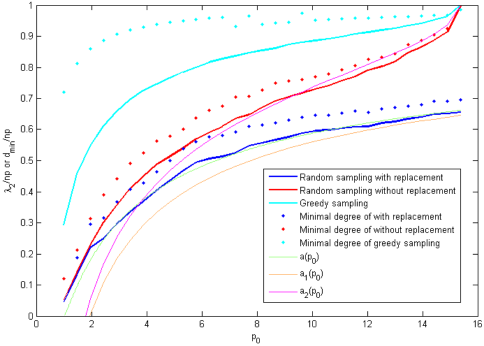


Figure: A comparison of the Fiedler value, minimal degree, and estimates $a(p_0)$, $a_1(p_0)$, and $a_2(p_0)$ for graphs generated via random sampling with/without replacement and greedy sampling at $n = 64$.

Active Sampling [Xu-Xiong-Chen-Huang-Y. AAAI'18]

- *Fisher Information Maximization*: Greedy sampling above, unsupervised
- *Bayesian Information Maximization*: supervised sampling
  - closed-form online formula based on Sherman-Morrison-Woodbury
  - faster and more accurate sampling scheme in literature

Table 2: Average running cost (s) of 100 runs on three real-world datasets.

Method	**Our supervised method**	**Crowd-BT**
VQA dataset	18	600
IQA dataset	12	480
Reading level dataset	120	4200

Figure: Note: Crowd-BT is proposed by Chen et al. 2013

Supervised active sampling is more accurate

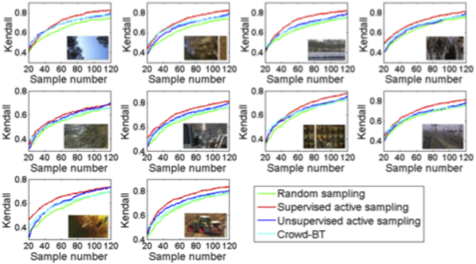
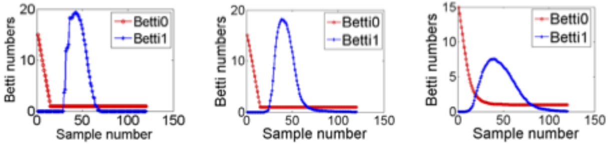


Figure 5: Experimental results of four sampling schemes for 10 reference videos in LIVE database.

Both supervised and unsupervised sampling reduce the chance of ranking chaos!



(a) Unsupervised. (b) Supervised. (c) Random.

Figure 2: Average Betti numbers for three sampling schemes.

Applications of Hodge Decomposition

- Boundary Value Problem (Schwarz, Chorin-Marsden'92)
- Computer vision
  - Optical flow decomposition and regularization (Yuan-Schnörr-Steidl'2008, etc.)
  - Retinex theory and shade-removal (Ma-Morel-Osher-Chien'2011)
  - Relative attributes (Fu-Xiang-Y. et al. 2014)
- Sensor Network coverage (Jadbabai et al.'10)
- Statistical Ranking or Preference Aggregation (Jiang-Lim-Y.-Ye'2011, etc.)
- Decomposition of Finite Games (Candogan-Menache-Ozdaglar-Parrilo'2011)

From Single Utility to Multiple Utilities

STRATEGIES	B Cooperate	B Defect
A Cooperate	(3,3)	(0,5)
A Defect	(5,0)	(1,1)

Prisoner's dilemma in Game Theory, (Flood-Dresher-Tucker 1950)

Voter 1	Voter 2	Voter 3	Voter 4
A>B>C	B>C>A	C>A>B	…

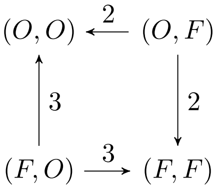
Voting theory and social choice

- Condorcet (1785), Borda (1700s)
- Kenneth Arrow (1972 Nobel Memorial Prize in Economics)
- Amartya Sen (1998 Nobel Memorial Prize in Economics)

Multiple Utility Flows for Games

	O	F
O	3, 2	0, 0
F	0, 0	2, 3

(a) Battle of the sexes



Extension to multiplayer games: $G = (V, E)$

- $V = \{(x_1, \ldots, x_n) =: (x_i, x_{-i})\} = \prod_{i=1}^{n} S_i$, n person game;
- undirected edge: $\{(x_i, x_{-i}), (x_i', x_{-i})\} = E$
- each player has utility function $u_i(x_i, x_{-i})$;
- Edge flow (1-form): $u_i(x_i, x_{-i}) - u_i(x_i', x_{-i})$

Nash and Correlated Equilibrium

$\pi(x_i, x_{-i})$, a joint distribution tensor on $\prod_i S_i$, satisfies $\forall x_i, x_i'$,

$$\sum_{x_{-i}} \pi(x_i, x_{-i})(u_i(x_i, x_{-i}) - u_i(x_i', x_{-i})) \geq 0,$$

i.e. expected flow ($\mathbb{E}[\cdot|x_i]$) is nonnegative. Then,

- tensor π is a correlated equilibrium (CE, Aumann 1974);
- if π is a rank-one tensor,

$$\pi(x) = \prod_i \mu(x_i),$$

  then it is a Nash equilibrium (NE, Nash 1951);
- pure Nash-equilibria are sinks;
- fully decided by the edge flow data.

What is a correct notion of Equilibrium?

- Players are never independent in reality, e.g. Bayesian decision process (Aumman'87)
- Finding NE is NP-hard, e.g. solving polynomial equations (Sturmfels'02, Datta'03)
- Finding CE is linear programming, easy for graphical games (Papadimitriou-Roughgarden'08)
- Some natural learning processes (best-response) converges to CE (Foster-Vohra'97)

Another simplification: Graphical Games

- n-players live on a network of n-nodes
- player i utility only depends on its neighbor players $N(i)$ strategies
- correlated equilibria allows a concise representation with parameters linear to the size of the network (Kearns et al. 2001; 2003)

$$\pi(x) = \frac{1}{Z} \prod_{i=1}^{n} \psi_i(x_{N(i)})$$

- this is not rank-one, but low-order interaction
- reduce the complexity from $O(e^{2^n})$ to $O(ne^{2^d})$ $(d = \max_i |N(i)|)$
- polynomial algorithms for CE in *tree* and *chodal* graphs.

Hodge Decomposition of Finite Games

Theorem (Candogan-Menache-Ozdaglar-Parrilo,2011)

Every finite game admits a unique decomposition:

$$\textit{Potential Games} \oplus \textit{Harmonic Games} \oplus \textit{Neutral Games}$$

Furthermore:

- Shapley-Monderer Condition: Potential games $\equiv$ quadrangular-curl free
- Extending $G = (V, E)$ to complex by adding quadrangular cells, harmonic games can be further decomposed into (quadrangular) curl games

Bimatrix Games

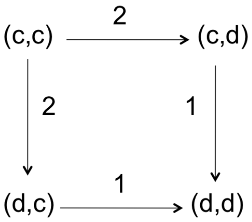
For bi-matrix game (A, B),

- potential game is decided by $((A + A')/2, (B + B')/2)$
- harmonic game is zero-sum $((A - A')/2, (B - B')/2)$
- Computation of Nash Equilibrium:
  - each of them is tractable
  - however direct sum is NP-hard
  - approximate potential game leads to approximate NE

Example: Hodge Decomposition of Prisoner's Dilemma

- Every game can be mapped to a flow preserving its Nash equilibrium
- Game flow = potential + harmonic

STRATEGIES	B Cooperate	B Defect
A Cooperate	(3,3)	(0,5)
A Defect	(5,0)	(1,1)



Note: Prisoner's dilemma is a potential game to its Nash equilibrium, not efficient! So we want new way for flow construction...

Candogan-Menache-Ozdaglar-Parrilo, 2010, Flows and Decompositions of Games: Harmonic and Potential Games, arXiv: 1004.2405v1, May 13, 2010.

Note: Shapley-Monderer Condition $\equiv$ Harmonic-free $\equiv$ quadrangular-curl free

What Does Hodge Decomposition Tell Us?

Does it suggest myopic greedy players might lead to

transient potential games + periodic equilibrium?

Basic Reference

- Jiang, Lim, Yao, and Ye, *Mathematical Programming*, 127(1): 203-244, 2011
- Candogan, Menache, Ozdaglar, and Parrilo, *Mathematics of Operational Research*, 36(3): 474-503, 2011

More reference

- Tran, N. M. Pairwise ranking: choice of method can produce arbitrarily different rank order. arXiv:1103.1110v1 [stat.ME], 2011
- Xu, Jiang, Yao, Huang, Yan, and Lin, *ACM Multimedia*, 2012
- Random graph sampling models: Erdös-Rényi and beyond
  - Xu, Jiang, Yao, Huang, Yan, and Lin, *IEEE T. Multimedia*, 2012
- Online algorithms
  - Xu, Huang, and Yao, *ACM Multimedia* 2012
- l_1-norm ranking
  - Osting, Darbon, and Osher, 2012
- Robust ranking:
  - Xu, Xiong, Huang, and Yao, *ACM Multimedia* 2013
  - Xu, Yang, Jiang, Cao, Huang, and Yao, *CVPR* 2019
- Mixed Effect/Personalized HodgeRank:
  - Xu, Xiong, Cao, and Yao, *ACM Multimedia* 2016
  - Xu, Xiong, Huang, Cao, and Yao, *IEEE T. PAMI, 2019*
  - Xu, Sun, Yang, Jiang, and Yao, *NeurIPS* 2019
  - Xu, Yang, Jiang, and Yao, *AAAI* 2020
- Active sampling
  - Osting, Brune, and Osher, *ICML* 2013
  - Osting, Xiong, Xu, and Yao, *ACHA* 2016
  - Xu, Xiong, Chen, Huang, and Yao, *AAAI* 2018

Summary

- New challenges from modern crowdsourced ranking data
- Hodge decomposition provides generalized Borda count in classical Social Choice
  - gradient flow, as generalized Borda count scores
  - curls/local cycles, as local inconsistency
  - harmonic flow, as global inconsistency or voting chaos

Such a decomposition has been seen in *computational fluid mechanics, computer vision, machine learning, sensor networks, and game theory*, etc. More are coming...