HodgeRank on Random Graphs

Yuan Yao

School of Mathematical Sciences
Peking University

July 15, 2011

1. Motivation
   - Crowdsourcing Ranking on Internet
2. HodgeRank on Random Graphs
   - HodgeRank on Graphs
   - Random Graph Models
   - Online Learning Algorithms
3. Application
   - Subjective Video Quality Evaluation
4. Discussions

Crowdsourcing Ranking on Internet

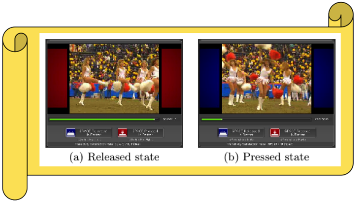


Figure: Start from a movie – *The Social Network*

Mean Opinion Score

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

widely used for evaluation of videos, as well books and movies, etc., but

- Ambiguity in definition of the scale;
- Difficult to verify whether a participant gives false ratings either intentionally or carelessly.

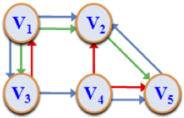
Paired Comparisons

- Individual decision process in paired comparison is simpler than in the typical MOS test, as the five-scale rating is reduced to a dichotomous choice;
- But the paired comparison methodology leaves a heavier burden on participants with a larger number $\binom{n}{2}$ of comparisons
- Moreover, raters and item pairs enter the system in a dynamic and random way;

Here we introduce:

Hodge Decomposition on Random Graphs
for paired comparisons

Pairwise Ranking Graphs



On a graph $G = (V, E)$,

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

- α for raters
- ω_{ij}^{α} is an indicator or confidence weight
- Y_{ij}^{α} is 1 if rater α prefers i to j and -1 otherwise

Equivalently, in weighted Least Square

$$\min_{s \in \mathbb{R}^{|V|}} \sum_{\{i,j\} \in E} \omega_{ij}(s_i - s_j - \hat{Y}_{ij})^2,$$

where

- $\hat{Y}_{ij} = (\sum_\alpha \omega_{ij}^\alpha Y_{ij}^\alpha)/(\sum_\alpha \omega_{ij}^\alpha)$, skew-symmetric matrix
- $\omega_{ij} = \sum_\alpha \omega_{ij}^\alpha$
- Inner product induced on R^E, $\langle u, v \rangle_\omega = \sum u_{ij} v_{ij} \omega_{ij}$ where u, v skew-symmetric

Note: NP-hard Kemeny Optimization, or Minimimum-Feedback-Arc-Set:

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

Linear Models in Statistics

Let π_{ij} be the probability that i is preferred to j. The family of linear models assumes that

$$\pi_{ij} = \Phi(s_i - s_j)$$

for some symmetric cumulated distributed function Φ. Reversely, given an observation $\hat{\pi}$, define

$$\hat{Y}_{ij} = \Phi^{-1}(\hat{\pi}_{ij})$$

One would like $\hat{Y}_{ij} \approx \hat{s}_i - \hat{s}_j$ for some $\hat{s} : V \to \mathbb{R}$ (in least squares, e.g.).

Examples of Linear Models

1. *Uniform* model:

$$\hat{Y}_{ij} = 2\hat{\pi}_{ij} - 1. \tag{1}$$

2. *Bradley-Terry* model:

$$\hat{Y}_{ij} = \log \frac{\hat{\pi}_{ij}}{1 - \hat{\pi}_{ij}}. \tag{2}$$

3. *Thurstone-Mosteller* model:

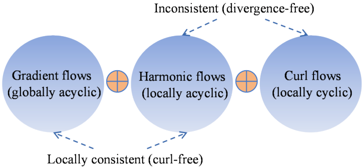
$$\hat{Y}_{ij} = \Phi^{-1}(\hat{\pi}_{ij}). \tag{3}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-x/[2\sigma^2(1-\rho)]^{1/2}}^{\infty} e^{-\frac{1}{2}t^2} dt.$$

4. *Angular transform* model:

$$\hat{Y}_{ij} = \arcsin(2\hat{\pi}_{ij} - 1). \tag{4}$$

HodgeRank on Graphs [Jiang-Lim-Y.-Ye 2011]



Every $\hat{Y}$ admits an orthogonal decomposition adapted to G,

$$\hat{Y} = \hat{Y}^{(1)} + \hat{Y}^{(2)} + \hat{Y}^{(3)}, \quad (5)$$

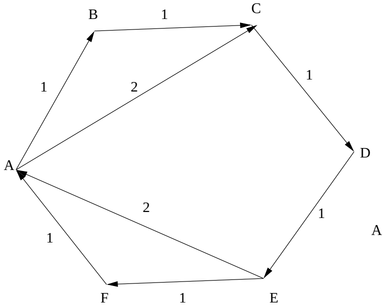
where

$$\hat{Y}^{(1)}_{ij} = \hat{s}_i - \hat{s}_j, \quad \text{for some } \hat{s} \in \mathbb{R}^V, \quad (6)$$

$$\hat{Y}^{(2)}_{ij} + \hat{Y}^{(2)}_{jk} + \hat{Y}^{(2)}_{ki} = 0, \text{ for each } \{i, j, k\} \in T, \quad (7)$$

$$\sum_{j \sim i} \omega_{ij} \hat{Y}^{(2)}_{ij} = 0, \text{ for each } i \in V. \quad (8)$$

Harmonic and Triangular Curl



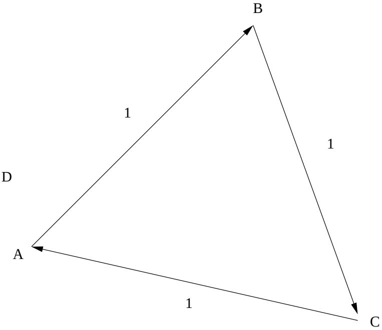


Figure: Left: example of $\hat{Y}^{(2)}$, harmonic; Right: example of $\hat{Y}^{(3)}$, curl.

Global Rating Score

The minimal norm least square solution $\hat{s}$ satisfies the normal eq.

$$\Delta_0 \hat{s} = \delta_0^* \hat{Y}, \tag{9}$$

where

- $\Delta_0 = \delta_0^* \cdot \delta_0$ is the unnormalized graph Laplacian defined by $(\Delta_0)_{ii} = \sum_{j \sim i} \omega_{ij}$ and $(\Delta_0)_{ij} = -\omega_{ij}$
- $\delta_0 : R^V \to R^E$ defined by $(\delta_0 v)(i,j) = v_i - v_j$
- $\delta_0^* = \delta_0^T W : R^E \to R^V$, $W = \mathrm{diag}(\omega_{ij})$, the adjoint of δ_0
- Spielman-Teng, Koutis-Miller-Peng et al. give provable almost-linear algorithms with suitable preconditioners

Local vs. Global Inconsistencies

Residues $\hat{Y}^{(2)}$ and $\hat{Y}^{(3)}$ accounts for inconsistencies, in different nature, which can be used to analyze rater's credibility or videos' confusion level .

- Define a 3-clique complex $\chi_G = (V, E, T)$ where
  - T collects all 3-cliques (complete subgraphs) $\{i, j, k\}$
- $\hat{Y}^{(3)}$, the local inconsistency, triangular curls
  - $\hat{Y}^{(3)}_{ij} + \hat{Y}^{(3)}_{jk} + \hat{Y}^{(3)}_{ki} \neq 0$, $\{i, j, k\} \in T$
- $\hat{Y}^{(2)}$, the global inconsistency, harmonic ranking
  - $\hat{Y}^{(2)}$ vanishes if 1-homology of χ_G vanishes
  - harmonic ranking is a circular coordinate and generally non-sparse $\Rightarrow$ fixed tournament issue

1-D Hodge Laplacian

- Define 1-coboundary map

$$\begin{aligned} \delta_1 \quad : \quad & \mathfrak{sl}(E) \subset \mathbb{R}^{V\times V} \to \mathbb{R}^{V\times V\times V} \\ & X \mapsto \pm(X_{ij} + X_{jk} + X_{ki})_{ijk} \end{aligned}$$

  where $\mathfrak{sl}(E)$ is skew-symmetric matrix on E.

- δ_1^* is the adjoint of δ_1.
- Define 1-Laplacian

$$\Delta_1 = \delta_0 \circ \delta_0^* + \delta_1^* \circ \delta_1$$

- $\dim(\ker \Delta_1) = \beta_1$
- $\hat{Y}^{(2)} = \mathrm{proj}_{\ker \Delta_1} \hat{Y}$

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: $P(i,j) = P(edge) = \int_0^1 \int_0^1 h(u,v)dudv =: p$

Why Topology?

To get a faithful ranking, two topological conditions important:

- Connectivity: G is connected, then an unique global ranking is possible;
- Loop-free: χ_G is loop-free, if one would like to avoid the fixed-tournament issue when Harmonic ranking is large.

Persistent Homology: online algorithm for topological change of evolutionary graphs

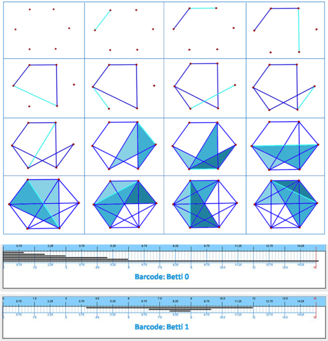
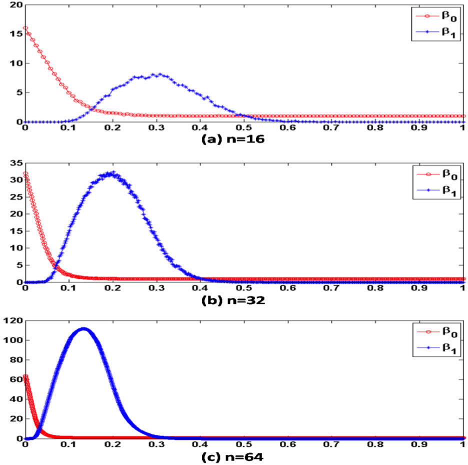


Figure: Persistent Homology Barcodes

- vertice, edges, and triangles etc. sequentially added
- online update of homology
- $O(m)$ for surface embeddable complex; and $O(m^3)$ in general (m number of simplex)

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.

An Intuition from Random Matrix Theory

Concentration of eigenvalues (Chung-Radcliffe 2011)

$$|\lambda_i(\tilde{\Delta}_0) - \lambda_i(\bar{\Delta}_0)| \leq O\left(\sqrt{np \log \frac{n}{\delta}}\right)$$

where

$$\bar{\Delta}_0(i,j) = npI_n - pee^T = \begin{cases} -p, & i \neq j \\ (n-1)p, & i = j \end{cases}$$

has one eigenvalue 0, and one eigenvalue np of multiplicity $n-1$

- $p >> n^{-1} \log n$, almost always large eigenvalues $np = \Omega(1)$;
- $p << n^{-1-\epsilon}$, almost always small eigenvalues $np = o(1)$;

1-Laplacian Splits

$$\tilde{\Delta}_1^{(l)}(ij,kl) = \delta_0 \circ \delta_0^* = \begin{cases} 2X_{ij} \to 2p, & \{i,j\} = \{k,l\} \\ \xi_{ij,kl}^{(l)} X_{ij} X_{jk} \to \xi_{ij,kl}^{(l)} p^2, & \text{otherwise} \end{cases}$$

where lower-coincidence number $\xi_{ij,kl}^{(l)} = \pm 1$ if $|\{i,j\} \cap \{k,l\}| = 1$ and 0 otherwise.

$$\tilde{\Delta}_1^{(u)}(ij,kl) = \delta_1^* \circ \delta_1 = \begin{cases} \sum_{ij\tau \in T} X_{ij} X_{j\tau} X_{\tau i} \to \frac{(np)(np^2)^n}{\log np^2}, & ij = kl \\ \xi_{ij,kl}^{(u)} X_{ij} X_{jk} X_{ki} \to \xi_{ij,kl}^{(u)} p^3, & \text{otherwise} \end{cases}$$

where upper-coincidence number $\xi_{ij,kl}^{(u)} = \pm 1$ if $|\{i,j\} \cup \{k,l\}| = 3$ and 0 otherwise.

- Forman (2003): $Ric_{\bar{\Delta}_1}(\text{ij}) =$ diagonal - sum of abs(off-diag)
- $p << n^{-1}$ or $p >> n^{-1/2}$, $\bar{\Delta}_1$ strongly diagonal dominant

Online HodgeRank as Stochastic Approximations

Robbins-Monro (1951) algorithm for $\bar{A}x = \bar{b}$

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

Now consider $\Delta_0 s = \delta_0^* \hat{Y}$, with new rating $Y_t(i_{t+1}, j_{t+1})$

$$\begin{aligned} s_{t+1}(i_{t+1}) &= s_t(i_{t+1}) - \gamma_t[s_t(i_{t+1}) - s_t(j_{t+1}) - Y_t(i_{t+1}, j_{t+1})] \\ s_{t+1}(j_{t+1}) &= s_t(j_{t+1}) + \gamma_t[s_t(i_{t+1}) - s_t(j_{t+1}) - Y_t(i_{t+1}, j_{t+1})] \end{aligned}$$

Note:

- updates only occur locally on edge $\{i_{t+1}, j_{t+1}\}$
- initial choice: $s_0 = 0$ or any vector $\sum_i s_0(i) = 0$
- step size (Smale-Yao 2006, Ying-Pontil 2007, etc.)
  - $\gamma_t = (t + c)^{-\theta}$ $(\theta \in (0, 1])$
  - $\gamma_t = const(T)$, .e.g. $1/T$ where T is total sample size

Averaging Process (Ruppert 1988; Y. 2010)

A second stage averaging process, following s_{t+1} above

$$z_{t+1} = \frac{t}{t+1} z_t + \frac{1}{t+1} s_{t+1}$$

with $z_0 = s_0$.
Note:

- Averaging process speeds up convergence for various choices of γ_t
- One often choose $\gamma_t = c$ to track dynamics
- In this case, z_t converges to $\hat{s}$ (population solution), with probability $1 - \delta$, in the (optimal) rate

$$\|z_t - \hat{s}\| \leq O\left(t^{-1/2} \cdot \kappa(\Delta_0) \cdot \log^{1/2} \frac{1}{\delta} \right)$$

Data Description

- Dataset: LIVE Database
- 10 different reference videos and 15 distorted versions of each reference, for a total of 160 videos.
- 32 rounds of complete comparisons are collected from 209 observers in lab. Because each round needs 1200 paired comparisons, the total number of comparisons for 32 rounds is $38400 = 32 \times 1200$.
- Note: we do not use the subjective scores in LIVE, we only borrow the video sources it provides.

Figure: Data collected from PKU junior undergraduates.

HodgeRank with Complete Data

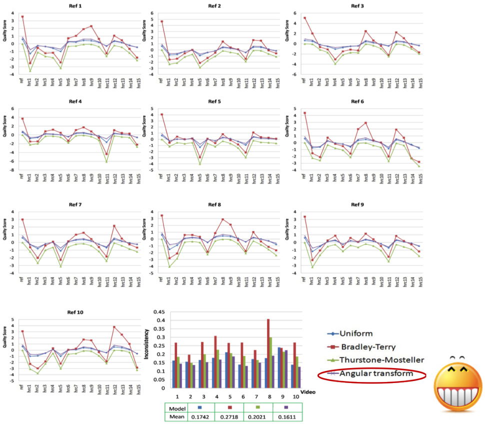


Figure: Angular Transform and Uniform models are the best two.

Globa/Harmonic and Local/Triangular Inconsistency

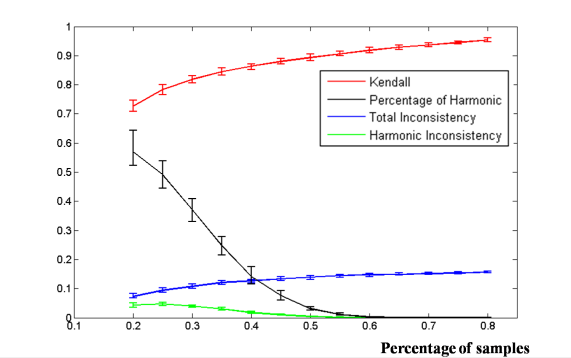


Figure: Harmonic inconsistency accounts for more than 50% total inconsistency before 25% edges, and rapidly drops to zero after 70% edges ($p \sim n^{-1/2}$)

Sampling Efficiency

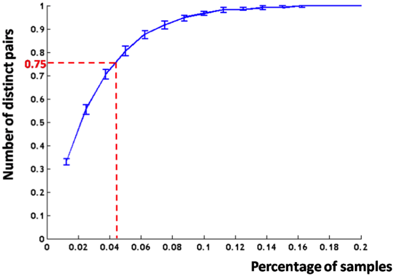
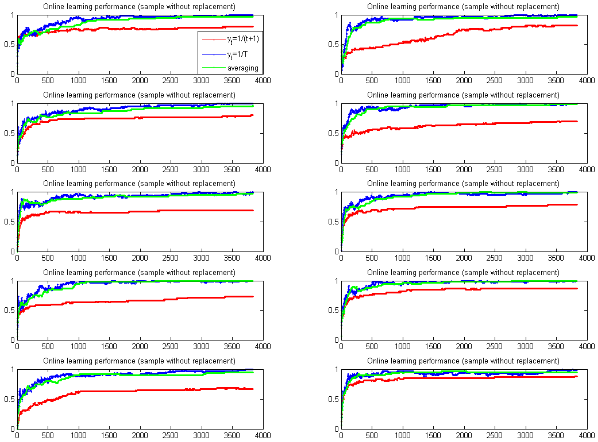


Table 3: Kendall's τ and inconsistency of of Exp-III.

	min	mean	max	std
Kendall's τ	0.8067	0.9337	0.9857	0.0415
Inconsistency	0.1623	0.2256	0.3777	0.0606

Convergence of Online Learning Algorithms



Discussions

- Erdös-Rényi random graphs give the simpliest sampling scheme, comparable to I.I.D. sampling in machine learning
- General random graphs (unlabeled) can use nonparametric models derived from generalized de Finetti's theorem (Bickel, Chen 2009)
- For computational concern, consider random graphs with small condition numbers, e.g. expanders
- For balancing concern, consider random k-regular graphs
- For top ranked videos, preference attachement models
- Markov sampling (Aldous, Vazirani 1990; Smale, Zhou 2007)
- Concentration inequalities with dependent random variables for high-dim Laplacians

Acknowledgement

- Reference: *Xu et al.* ACM Multimedia 2011, *to appear*.
- Experiments:
  - Qianqian Xu (Chinese Academy of Sciences)
  - Bowei Yan (Peking University)
- Discussions:
  - Tingting Jiang (Peking University)
  - Qingming Huang (Chinese Academy of Sciences)
  - Lek-Heng Lim (U Chicago)
  - Sayan Mukherjee (Duke)
  - Gunnar Carlsson (Stanford)
  - Steve Smale (City University of Hong Kong)
  - Shmuel Weinberger (U Chicago)
  - Yinyu Ye (Stanford)