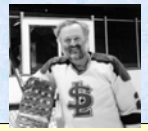
Rating the Ratings: A Comparison of Methods for Ranking College Ice Hockey Teams

Julia Palmateer, St. Lawrence University & The MathWorks
Robin Lock, St. Lawrence University



The problem:

- Assess methods for ranking teams that play a majority of their games within leagues that may differ in strength.

The approach:

- Use Monte Carlo simulation of many seasons to compare the effectiveness of different ratings methods to rank order teams with assumed "known" abilities.

WPct: Raw Winning Percentage

- Ties count as ½ win

RPIw: Ratings Percentage Index[1]

- Adjust for strength of opponents

$$RPIw = 0.3*WPct + 0.24*AvgOppWpct + 0.46*OppOppWPct$$

Or other weights, e.g. 25%, 21%, 54% for *RPIm*

KRACH: Ken's Rating for American College Hockey[2]

- Recursively-defined Bradley-Terry model

$$KRACH = K_{You} = \frac{WinRatio}{\sum_{Games} \frac{1}{K_{You} + K_{Opp}}} * \sum_{Games} \frac{K_{Opp}}{K_{You} + K_{Opp}}$$

Standardize so that WPct=0.500 ➔ Krach=100

The CHODR[3] simulation method:

When Team A plays Team B:

Goals for A ~ Poisson(λ_{AB})

where $$\lambda_{AB} = \frac{Off_A \times Def_B}{\mu} \cdot H^{\pm 1}$$

μ = average scoring rate, H = home ice advantage

The "true" rankings:

Assume a balanced schedule among all teams and use the CHODR Poisson scoring rates to compute a probability of winning vs. every other team. The **expected winning percentage** is

$$EWPct = \frac{\sum_{AllOpponents} P(Win) + \frac{1}{2} P(Tie)}{\#Opponents}$$

Rank teams by EWPct.

Hypothetical: Strong vs. Weak League

Use 12 teams in the Women's ECAC and divide into two leagues, Holly and Ivy. Assign team abilities so that

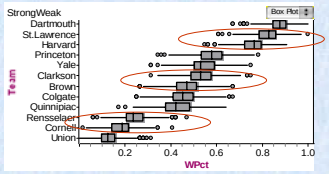
- The Ivy league is stronger overall than Holly
- Equivalent teams from both leagues can be paired.
- Average scoring rate μ = 3.0, Home ice H=1.05
- Schedule: 32 games in a season for each team
  - 4 games vs. each league opponent
  - 2 games vs. each team in the other league

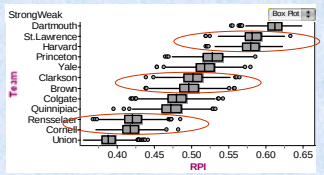
Hypothetical Team Abilities

IVY	Off	Def	EWPct	HOLLY	Off	Def	EWPct
Dartmouth	4.8	1.7	.865	St. Lawrence	4.5	2.3	.764
Harvard	4.5	2.3	.764	Clarkson	3.0	3.0	.500
Princeton	2.6	2.0	.597	Colgate	2.4	2.9	.426
Yale	2.9	2.4	.574	Quinnipiac	2.0	2.6	.403
Brown	3.0	3.0	.500	Rensselaer	2.3	4.5	.236
Cornell	2.3	4.5	.236	Union	1.7	4.8	.135

A Typical Simulated Season

	W	L	T	WPct	RPI	Krach
Dartmouth	26	3	3	.859	.606	793
St. Lawrence	26	4	2	.844	.589	532
Harvard	23	6	3	.765	.580	448
Princeton	19	8	5	.672	.554	273
Colgate	16	13	3	.547	.506	106
Brown	16	15	1	.516	.511	111
Quinnipiac	15	14	3	.516	.497	91
Clarkson	14	17	1	.453	.480	67
Yale	11	18	3	.391	.476	69
Rensselaer	7	24	1	.234	.420	20
Cornell	5	27	0	.156	.411	17
Union	1	30	1	.047	.368	4





Results for 1000 simulated seasons

	Ranks	1st	2nd	3rd	4th	5th	AvgRnk	Higher
Wpct	Harvard	53	226	672	46	3	2.72	23.2%
	SLU	245	546	203	5	1	1.97	76.8%
RPI	Harvard	78	394	485	41	2	2.50	45.4%
	SLU	108	458	401	31	2	2.36	54.6%
KRACH	Harvard	74	420	467	38	1	2.47	47.6%
	SLU	136	403	408	48	5	2.38	52.4%

Avg.Rank	WPct	RPI	KRACH
Brown	7.22	6.67	6.49
Clarkson	5.78	6.33	6.50
%Clrk>Brwn	76.3%	56.7%	49.8%

Avg.Rank	WPct	RPI	KRACH
Cornell	11.02	10.60	10.58
Rensselaer	10.06	10.51	10.53
%Rnsl>Crnl	76.9%	53.2%	51.3%

Realistic Schedule: NCAA Women

Use the actual schedule from the 2006-7 season for 32 women's teams, with team abilities based on real game results and CHODR to simulate 1000 seasons.

League	Teams	Avg. Off.	Avg. Def	AvgEWpct	Avg. E(Rank)
WCHA	8	3.09	2.25	0.590	13.0
ECAC	12	2.67	2.76	0.497	16.2
CHA	4	2.70	2.80	0.473	17.9
Hockey East	8	2.36	3.01	0.428	19.8

Rank	Team	Off	Def	EWPct
1	Wisconsin	4.1	1.0	0.871
2	Harvard	4.2	1.4	0.821
3	Dartmouth	4.2	1.5	0.807
4	Mercyhurst	4.1	1.5	0.800
5.5	Minnesota-Duluth	3.8	1.4	0.793
5.5	New_Hampshire	3.8	1.4	0.793
7	St._Lawrence	4.1	1.9	0.744
8	Boston_College	3.7	2.0	0.696
9	Minnesota	3.7	2.1	0.682
10	Ohio_State	3.3	2.0	0.657
11	Colgate	2.5	1.9	0.580
12	Yale	3.0	2.4	0.566
13	Princeton	2.6	2.1	0.561
14	Minnesota_State	3.4	2.9	0.546
15	Clarkson	2.4	2.1	0.534
16	St._Cloud_State	3.0	2.7	0.526
17	Providence	2.6	2.4	0.517
18	Connecticut	2.2	2.3	0.476
19	Wayne_State	3.0	3.5	0.432
20	Bemidji_State	2.1	2.6	0.421
21.5	Niagara	2.1	2.7	0.408
21.5	Rensselaer	2.1	2.7	0.408
23	Boston_University	2.0	2.8	0.381
24	Brown	2.3	3.3	0.367
25	Quinnipiac	2.4	4.1	0.305
26	Maine	1.9	3.6	0.284
27	Robert_Morris	1.6	3.5	0.251
28.5	North_Dakota	1.3	3.3	0.224
28.5	Cornell	1.3	3.3	0.224
30	Northeastern	2.0	4.8	0.209
31	Vermont	0.7	4.8	0.067
32	Union	0.9	6.4	0.048

Which of equivalent teams is ranked higher?

	UMD	UNH
WinPct	196	804
RPIw	409	591
RPIm	474	526
KRACH	499	501

	Niag	Rens
WinPct	466	534
RPIw	456	544
RPIm	454	546
KRACH	468	532

	UND	Cor
WinPct	124	876
RPIw	243	718
RPIm	258	692
KRACH	435	513

Wrong team "wins"?

	SCS	Prov
WinPct	317	683
RPIw	383	617
RPIm	433	567
KRACH	522	478

	MnS	Clk
WinPct	164	834
RPIw	282	718
RPIm	308	742
KRACH	487	513

	UND	NE
WinPct	69	931
RPIw	386	614
RPIm	431	569
KRACH	552	448

Observations

- (Obviously) **WPct** is strongly biased to favor teams playing a weaker schedule.
- **RPI** still gives some advantage to teams from the weaker league, more so for the women's weights than the men's.
- **KRACH** does the best job of balancing equivalent teams from different strength leagues.
- Under each rating method a strong team (e.g. St. Lawrence in the hypothetical simulation) has a better chance of finishing first by playing in a weaker league, than an equivalent team (Harvard) that plays in a league with other strong teams.

References

1. John Whelan's RPI page: http://slack.net/~whelan/tbrw/2007/rpi.shtml
2. USCHO's KRACH FAQ: http://www.uscho.com/FAQs/?data=krach
3. Lock, R. and Danehy, T. "CHODR - Using Statistics to Predict College Hockey", *STATS*, Vol. 13, pp. 10-14