How to compare apples with oranges?

Balázs R. Sziklai

KRTK

Corvinus University of Budapest

sziklai.balazs@krtk.mta.hu

2022. 11. 22.

Acknowledgement

Supported by the ÚNKP-22-5 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

Sum of Ranking Differences (SRD)

- SRD is a novel statistical test that compares solution through a reference (Héberger, 2010).
- SRD is rapidly gaining popularity in science. Applications ranging from machine learning, through multi-criteria decision-making and pharmacology, to political science, and even sports.

The model

- M is an $n \times m$ matrix,
- where the first $m-1$ columns represent the different models,
- while the rows represent the measured variables (properties);
- the last column contains the reference values.

Example input

Districts	Mom. Inv. ($\beta = -0.5$)	Mom. Inv. ($\beta = 1$)	Mom. Inv. ($\beta = 2$)	Lee-Sallee	Reock	Polsby-Popper	Length-to-width	Reference
Arkansas 1st	0.936	0.810	0.584	0.721	0.396	0.144	0.924	0.645
Arkansas 2nd	0.924	0.640	0.301	0.582	0.311	0.221	0.693	0.524
Arkansas 3rd	0.940	0.698	0.365	0.619	0.328	0.327	0.824	0.586
Arkansas 4th	0.947	0.753	0.474	0.617	0.394	0.260	0.292	0.534
Iowa 1st	0.944	0.790	0.527	0.655	0.388	0.403	0.980	0.670
Iowa 2nd	0.895	0.504	0.170	0.483	0.208	0.255	0.720	0.462
Iowa 3rd	0.881	0.544	0.224	0.445	0.254	0.302	0.025	0.382
Iowa 4th	0.948	0.758	0.483	0.610	0.428	0.468	0.549	0.606
Iowa 5th	0.945	0.729	0.399	0.654	0.273	0.323	0.418	0.534
Kansas 1st	0.950	0.734	0.430	0.790	0.387	0.431	0.000	0.532
Kansas 2nd	0.854	0.577	0.298	0.439	0.355	0.230	0.353	0.443
Kansas 3rd	0.910	0.743	0.472	0.619	0.389	0.355	0.942	0.633
Kansas 4th	0.923	0.655	0.332	0.549	0.346	0.467	0.343	0.516

SRD step by step

1. Defining the reference (data fusion)
2. Rank transformation of the input
3. Computing the SRD values for the solutions (distance from the reference)
4. Validation (CRRN and cross-validation)

1. Reference values

SRD requires a reference value for each object. In some cases, justified reference values are available (prescriptions, earlier measurements). In the absence of a known gold standard, these reference values have to be extracted from the data. This step is called the *data fusion*. Most common reference values are:

- Known gold standard
- Average (arithmetic mean)
- Minimum/maximum
- Median

2. Converting the data matrix

- We create a ranking matrix by replacing each value in the column by its rank.
- That is, for each column (including the reference) take the smallest value in the column and replace it with '1', take the second smallest value and replace it with '2', and so on. Finally, the last remaining value, which was the largest of the original column values, is replaced by 'n'.
- Ties in column vectors are resolved by giving the same rank to cells with the same value: the arithmetic mean of the ranks.

3. Computing the SRD values

- We calculate the (absolute) ranking differences between the reference and solution vector coordinates and sum them up.
- The SRD values are, in fact, city block (Manhattan) distances, and they rank the solutions.
- The smaller the SRD value the closer the solution is to the benchmark, *i.e.* the better.
- The mutual proximity of SRD values indicates the specific grouping of variables.

4. Validation

- To remain comparable within various data sets (and different number of rows) the normalized SRD values are calculated.
- The permutation test (also called randomization test, denoted by CRRN = comparison of ranks with random numbers) shows whether the rankings are comparable with a ranking taken at random or they are different from it significantly.
- The second validation option is called cross-validation, and assigns uncertainties to the SRD values. Leave-one-out cross-validation is applied if the number of rows is less than 14. Leave-many out cross-validation is applied for larger number of rows in the input matrix.

Fair representation

The problem

In most democratic countries, some or all members of the Parliament are elected directly by the voters in electoral districts or (single-member) constituencies. For practical considerations these constituencies are embedded in the countries' existing administrative units, such as states or counties. To ensure equal representation, states are allotted seats in proportion to their populations. There are two related problems:

- How to distribute the seats among the administrative regions? (apportionment)
- How to design the constituencies, that is, how to draw the boundaries? (gerrymandering)

Apportionment in fair representation

Difficulties

- The division problems stems from the fact that fractional seats cannot be allocated (indivisible objects).
- The sizes of the constituencies should be roughly the same. Under ideal circumstances, every constituency contains the same number of voters.
- Constituency boundaries may be affected by the geography of the region, by administrative or historic boundary lines, or because of the concentration of a specific national minority.

Gerry-salamander

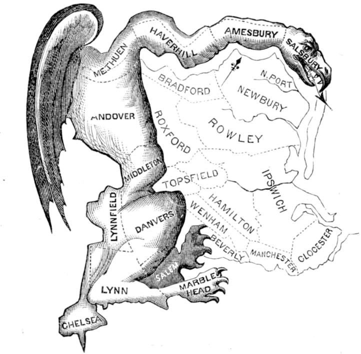


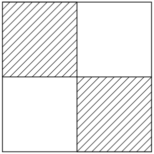
Figure: Elbridge Gerry választókörzete

Gerrymandering: the manipulation of the constituency boundaries to favor one political party

Example - Gerrymandering

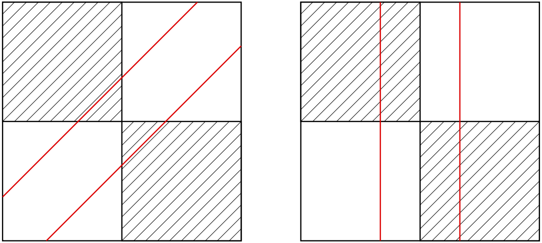
Exercise

Suppose two parties compete in a region where there are three electoral districts. The supporters of the two parties are equal in numbers but concentrate on different areas of the region. Let us assume that population are spread homogeneously on the map as follows

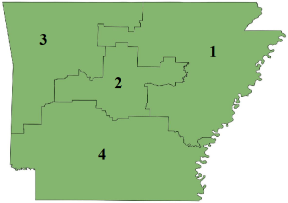


Create districts that favor one of the parties! How can we achieve fair representation?

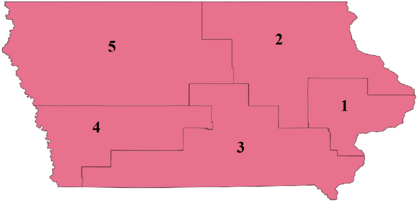
Solution



Compactness - Gerrymandering



Arkansas



Iowa

Kansas

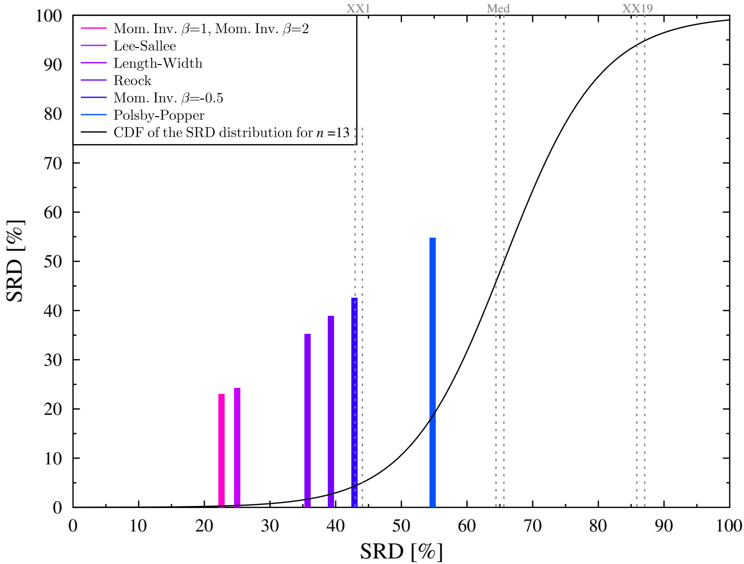
Compactness measures

Districts	Mom. Inv. ($\beta = -0.5$)	Mom. Inv. ($\beta = 1$)	Mom. Inv. ($\beta = 2$)	Lee-Sallee	Reock	Polsby-Popper	Length-to-width	Reference
Arkansas 1st	0.936	0.810	0.584	0.721	0.396	0.144	0.924	0.645
Arkansas 2nd	0.924	0.640	0.301	0.582	0.311	0.221	0.693	0.524
Arkansas 3rd	0.940	0.698	0.365	0.619	0.328	0.327	0.824	0.586
Arkansas 4th	0.947	0.753	0.474	0.617	0.394	0.260	0.292	0.534
Iowa 1st	0.944	0.790	0.527	0.655	0.388	0.403	0.980	0.670
Iowa 2nd	0.895	0.504	0.170	0.483	0.208	0.255	0.720	0.462
Iowa 3rd	0.881	0.544	0.224	0.445	0.254	0.302	0.025	0.382
Iowa 4th	0.948	0.758	0.483	0.610	0.428	0.468	0.549	0.606
Iowa 5th	0.945	0.729	0.399	0.654	0.273	0.323	0.418	0.534
Kansas 1st	0.950	0.734	0.430	0.790	0.387	0.431	0.000	0.532
Kansas 2nd	0.854	0.577	0.298	0.439	0.355	0.230	0.353	0.443
Kansas 3rd	0.910	0.743	0.472	0.619	0.389	0.355	0.942	0.633
Kansas 4th	0.923	0.655	0.332	0.549	0.346	0.467	0.343	0.516

SRD calculation

Districts	Mom. Inv. ($\beta = -0.5$)	Mom. Inv. ($\beta = 1$)	Mom. Inv. ($\beta = 2$)	Lee-Sallee	Reock	Polsby-Popper	Length-to-width	Ref. ranking
Arkansas 1st	7	13	13	12	12	1	11	12
Arkansas 2nd	6	4	4	5	4	2	8	5
Arkansas 3rd	8	6	6	8	5	8	10	9
Arkansas 4th	11	10	10	7	11	5	3	7
Iowa 1st	9	12	12	11	9	10	13	13
Iowa 2nd	3	1	1	3	1	4	9	3
Iowa 3rd	2	2	2	2	2	6	2	1
Iowa 4th	12	11	11	6	13	13	7	10
Iowa 5th	10	7	7	10	3	7	6	8
Kansas 1st	13	8	8	13	8	11	1	6
Kansas 2nd	1	3	3	1	7	3	5	2
Kansas 3rd	4	9	9	9	10	9	12	11
Kansas 4th	5	5	5	4	6	12	4	4
SRD value	36	19	19	21	33	46	30	
SRD (norm)	0.428	0.226	0.226	0.250	0.393	0.548	0.3579	

CRRN - Gerrymandering

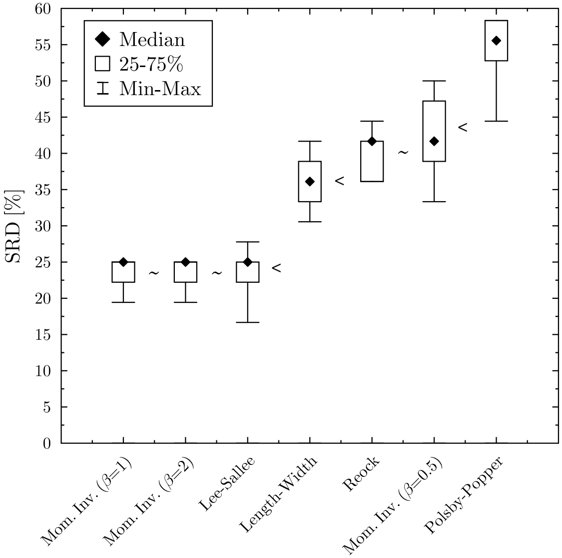


Cross-validation

We create k folds by leaving out some of the rows, then re-calculate the SRD values. The obtained scores are tested with the Wilcoxon signed rank test.

Districts	Mom. Inv. ($\beta = -0.5$)	Mom. Inv. ($\beta = 1$)	Mom. Inv. ($\beta = 2$)	Lee-Sallee	Reock	Polsby-Popper	Length-to-width	Ref.
Arkansas 1st	0.936	0.81	0.584	0.721	0.396	0.144	0.924	0.645
Arkansas 2nd	0.924	0.64	0.301	0.582	0.311	0.221	0.693	0.524
Arkansas 3rd	0.94	0.698	0.365	0.619	0.328	0.327	0.824	0.586
Arkansas 4th	0.947	0.753	0.474	0.617	0.394	0.26	0.292	0.534
Iowa 1st	0.944	0.79	0.527	0.655	0.388	0.403	0.98	0.67
Iowa 2nd	0.895	0.504	0.17	0.483	0.208	0.255	0.72	0.462
Iowa 3rd	0.881	0.544	0.224	0.445	0.254	0.302	0.025	0.382
Iowa 4th	0.948	0.758	0.483	0.61	0.428	0.468	0.549	0.606
Iowa 5th	0.945	0.729	0.399	0.654	0.273	0.323	0.418	0.534
Kansas 1st	0.95	0.734	0.43	0.79	0.387	0.431	0	0.532
Kansas 2nd	0.854	0.577	0.298	0.439	0.355	0.23	0.353	0.443
Kansas 3rd	0.91	0.743	0.472	0.619	0.389	0.355	0.942	0.633
Kansas 4th	0.923	0.655	0.332	0.549	0.346	0.467	0.343	0.516

Cross-validation



Case study I - Food laboratories

- Laboratories participate in a comparison program, where they have to determine some characteristics of a homogeneous sample under documented conditions.
- Polycyclic Aromatic Hydrocarbon contents in 16 edible oil samples were reported by each participating laboratory.
- Since the laboratories work with the same substances, **the laboratories should report the same measurements** (within a small statistical error).
- Reference values were provided by the European Union Reference Laboratory for PAHs in food (EU-RL-PAH).

Case study I - Food laboratories

Compound	L1	L4	L10	L11	Standard
5-Methylchrysene	1.20	1.01	1.30	1.13	**1.1**
Benzo[a]anthracene	2.40	2.48	2.30	2.72	**2.4**
Benzo[a]pyrene	2.90	3.06	2.80	3.13	**3.0**
Benzo[b]fluoranthene	5.20	5.52	5.40	5.66	**5.4**
Benzo[c]fluorene	2.20	1.91	1.30	1.83	**1.8**
Benzo[ghi]perylene	6.10	6.44	5.80	6.58	**6.2**
Benzo[j]fluoranthene	1.40	1.49	1.70	1.01	**1.4**
Benzo[k]fluoranthene	8.20	8.23	8.80	8.61	**8.2**
Chrysene	3.70	3.58	3.30	3.87	**3.4**
Cyclopenta[cd]pyrene	8.60	8.28	6.20	7.17	**7.7**
Dibenzo[a,e]pyrene	0.80	0.78	0.80	0.78	**1.0**
Dibenzo[a,h]anthracene	4.90	4.83	4.40	5.05	**3.8**
Dibenzo[a,h]pyrene	2.10	2.23	2.10	1.95	**2.5**
Dibenzo[a,i]pyrene	9.10	9.31	10.30	9.41	**9.8**
Dibenzo[a,l]pyrene	1.60	1.13	1.60	1.41	**1.5**
Indeno[1,2,3-cd]pyrene	3.40	3.82	3.80	3.81	**3.8**

Case study II - Chess

- Elo ratings of the participants of the Grand Swiss tournament of 2019
- Data: Pre- and post tournament Elo ratings and tournament performance
- Post tournament ratings are the best approximations for the current playing strength of the players, hence it is chosen as the reference.
- Preliminary ratings and tournament performances are two perturbations of different amplitude - **CV methods should be able to distinguish between the two.**

Case study II - Chess

Name	Country	Starting Rating	Performance	Finish rating
Magnus Carlsen	NOR	2876	2825	2870
Fabiano Caruana	USA	2812	2888	2822.2
Levon Aronian	ARM	2758	2833	2769.5
Alexander Grischuk	RUS	2759	2779	2761.7
Wesley So	USA	2767	2705	2758.5
Viswanathan Anand	IND	2765	2707	2757.1
Yangyi Yu	CHN	2763	2720	2756.8
Hikaru Nakamura	USA	2745	2803	2753.6
Sergey Karjakin	RUS	2760	2707	2752.8
Hao Wang	CHN	2726	2900	2750.7
Radoslaw Wojtaszek	POL	2748	2714	2743.1
Harikrishna Pentala	IND	2748	2698	2741
Nikita Vitiugov	RUS	2732	2792	2741
Vladislav Artemiev	RUS	2746	2632	2729.1
Peter Svidler	RUS	2729	2710	2726.2
...	...	...	...	...

Case study III - Influence maximization

- In the influence maximization problem we are trying to optimize a network diffusion (e.g. innovation spreading)
- In the original version we aim to find the k most influential persons, whose activation (by a marketing campaign) would incite the largest influence spread in the network.
- Here we try to assess the influence of groups instead of individuals. A real life example would be when a politician tries to decide which towns to visit in a campaign.
- question: which network centrality could predict the influence spread the best?

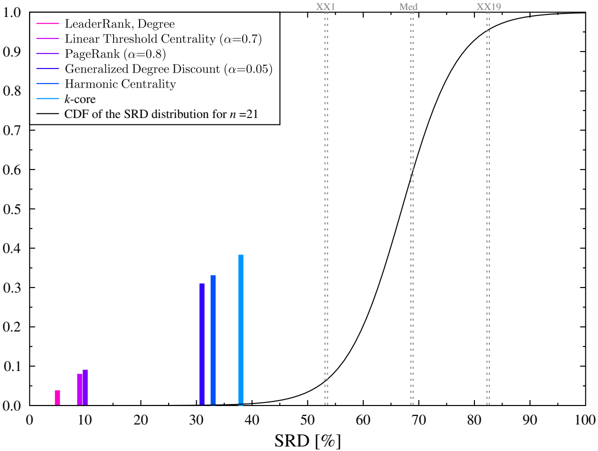
Case study III - Influence maximization

	Degree	Harmonic	PageRank	GDD(0.05)	k-core	LTC(0.7)	Shapley(G1)	TC	Avg. Spread
S1	19.686	0.2197	3.6226	32.924	10.152	22.522	0.9846	0.2853	3.271
S2	20.132	0.2210	3.6639	33.660	10.532	22.964	0.9784	0.2971	3.351
S3	19.500	0.2205	3.6198	32.976	10.248	22.362	0.9833	0.2887	3.275
S4	20.058	0.2197	3.6983	32.947	10.078	23.014	1.0170	0.2875	3.337
S5	19.664	0.2199	3.6625	33.037	10.226	22.570	1.0114	0.2879	3.294
S6	18.300	0.2193	3.4616	32.580	10.100	20.970	0.9470	0.2742	3.074
S7	20.972	0.2212	3.7904	34.108	10.606	23.932	1.0245	0.3011	3.461
S8	20.848	0.2215	3.8019	35.055	10.838	23.942	1.0265	0.3010	3.442
S9	18.948	0.2195	3.5251	32.467	10.036	21.768	0.9607	0.2821	3.162
S10	19.538	0.2210	3.6277	33.875	10.560	22.334	0.9836	0.2938	3.265
S11	19.486	0.2196	3.5936	33.431	10.334	22.254	0.9698	0.2908	3.239
S12	20.284	0.2213	3.7126	33.744	10.498	23.136	0.9973	0.2985	3.379
S13	20.166	0.2195	3.7371	33.069	10.204	23.148	1.0296	0.2962	3.355
S14	20.076	0.2193	3.7118	33.288	10.058	23.000	1.0267	0.2831	3.333
S15	20.268	0.2215	3.7231	33.826	10.500	23.184	1.0070	0.2990	3.367
S16	20.658	0.2217	3.8264	34.869	10.786	23.680	1.0436	0.3040	3.424
S17	20.764	0.2206	3.7817	33.879	10.446	23.678	1.0211	0.3045	3.455
S18	20.544	0.2215	3.7619	34.239	10.602	23.438	1.0093	0.2983	3.413
S19	19.662	0.2201	3.6355	33.450	10.382	22.472	0.9882	0.2971	3.281
S20	19.072	0.2189	3.5767	32.822	9.964	21.864	0.9807	0.2759	3.199
S21	19.874	0.2212	3.6521	34.361	10.672	22.836	0.9790	0.2936	3.293

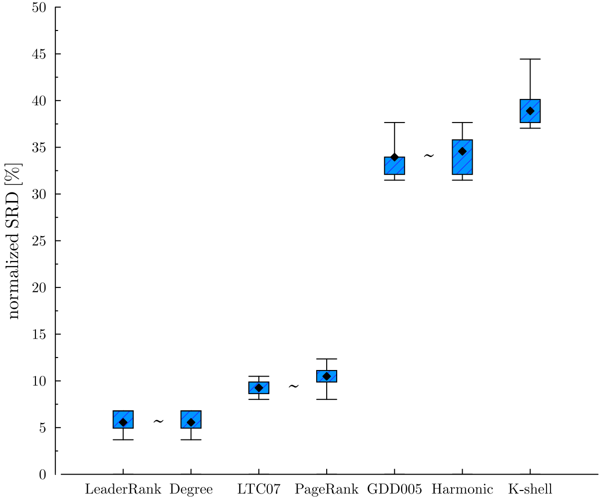
Case study III - Influence maximization

	Degree	Harmonic	PageRank	GDD(0.05)	k-core	LTC(0.7)	Shapley(G1)	TC	Avg. Spread
S1	9	7	6	4	6	8	9	5	6.5
S2	13	13	11	12	15	11	4	14	13.5
S3	5	11	5	6	9	6	7	8	6.5
S4	11	8	12	5	4	13	15	6	11.5
S5	8	9	10	7	8	9	14	7	9.5
S6	1	2	1	2	5	1	1	1	1
S7	21	16	19	17	18	20	17	19	21
S8	20	19	20	21	21	21	18	18	19
S9	2	4	2	1	2	2	2	3	2
S10	6	14	7	15	16	5	8	11	5
S11	4	6	4	10	10	4	3	9	4
S12	16	17	14	13	13	14	11	16	16
S13	14	5	16	8	7	15	20	12	13.5
S14	12	3	13	9	3	12	19	4	11.5
S15	15	20	15	14	14	16	12	17	15
S16	18	21	21	20	20	19	21	20	18
S17	19	12	18	16	12	18	16	21	20
S18	17	18	17	18	17	17	13	15	17
S19	7	10	8	11	11	7	10	13	8
S20	3	1	3	3	1	3	6	2	3
S21	10	15	9	19	19	10	5	10	9.5
SRD	12	73	22	69	83	19	72	49	0
nSRD	0.055	0.332	0.100	0.314	0.377	0.086	0.327	0.223	0.000

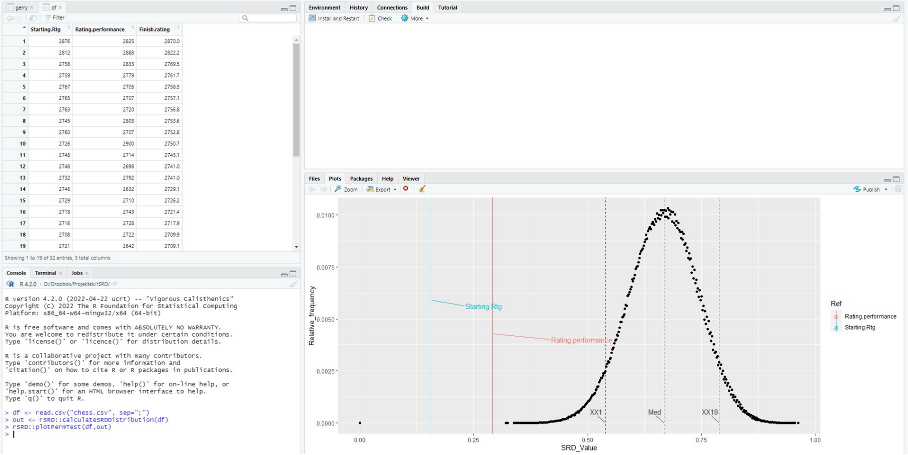
Case study III - Influence maximization



Case study III - Influence maximization



rSRD package downloadable soon from CRAN



Thank you for your attention!