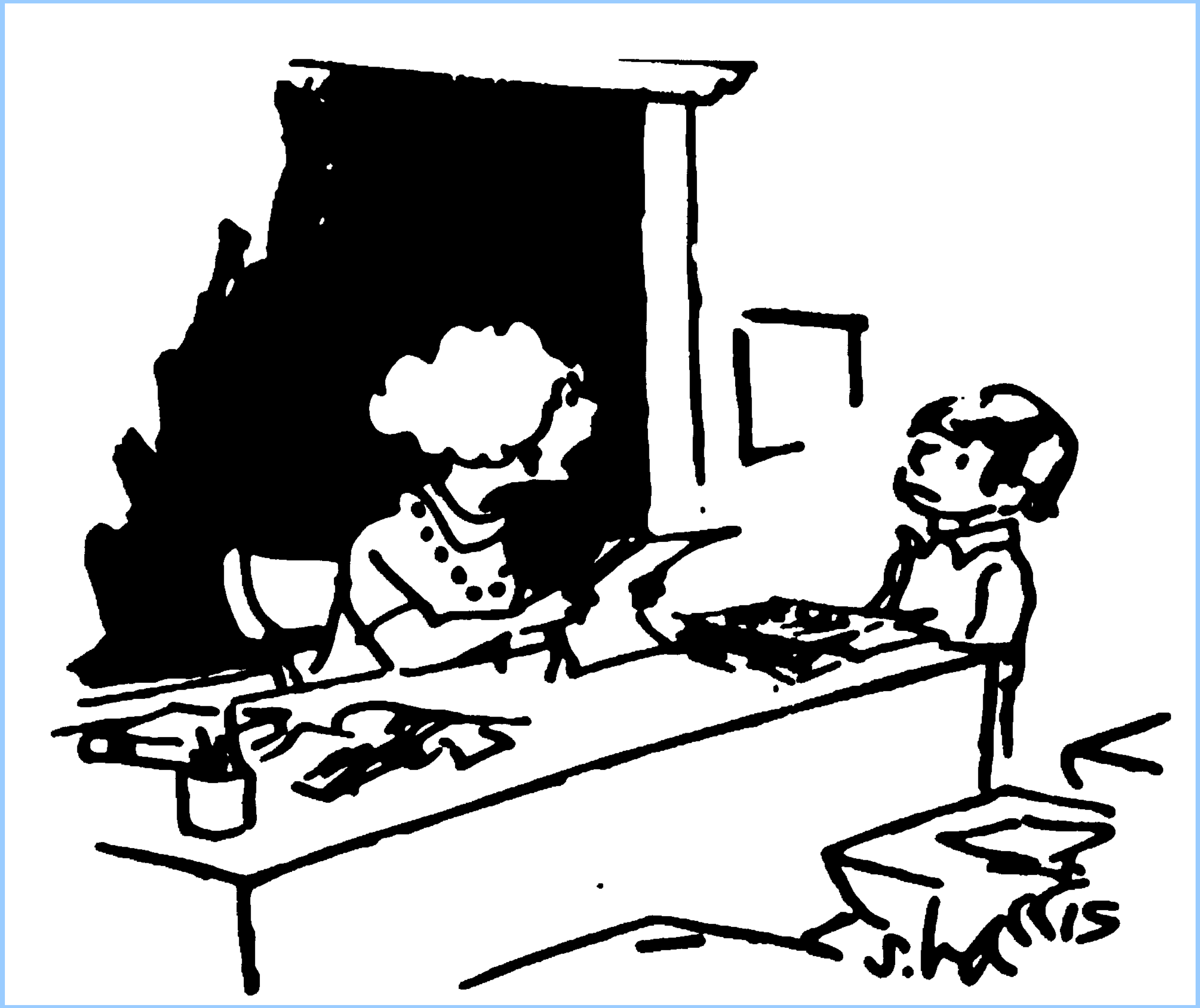


This report is very disappointing.
What kind of software are you using?

Stone-Age Space-Age Syndrome

	Stone-age **data**	Space-age **data**
Stone-age **analysis**	⊕	⊕
Space-age **analysis**		⊕

Multiple Indicators
Partial Orderings
and
Ranking and Prioritization
without Combining Indicators

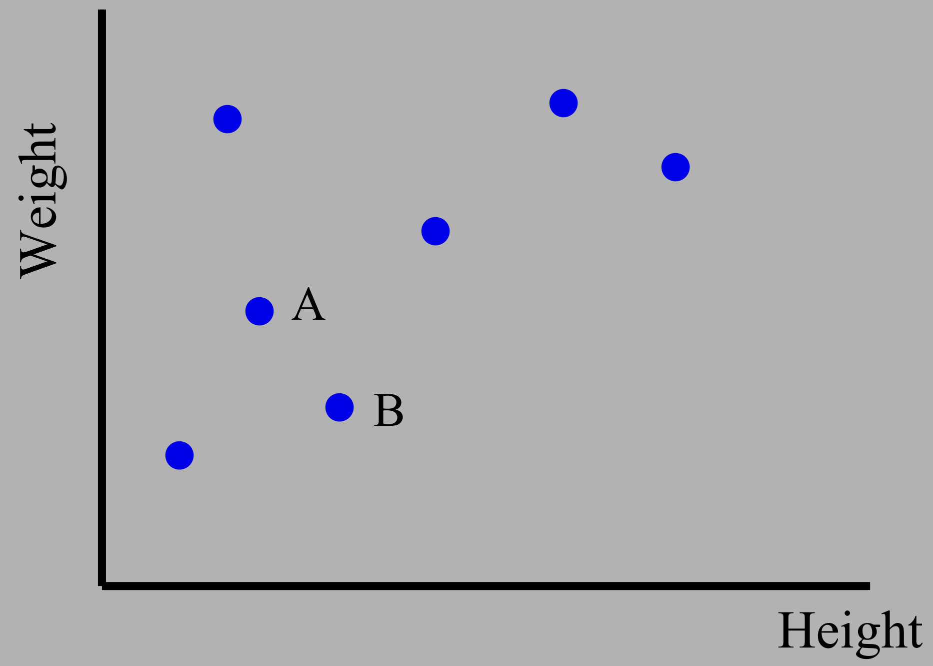
G. P. Patil
C. Taillie

Multiple Indicators

Two indicators of a person's **size**:

$$I_1 = \text{Height}$$
$$I_2 = \text{Weight}$$

Scatter Plot:



Is A bigger than B ?

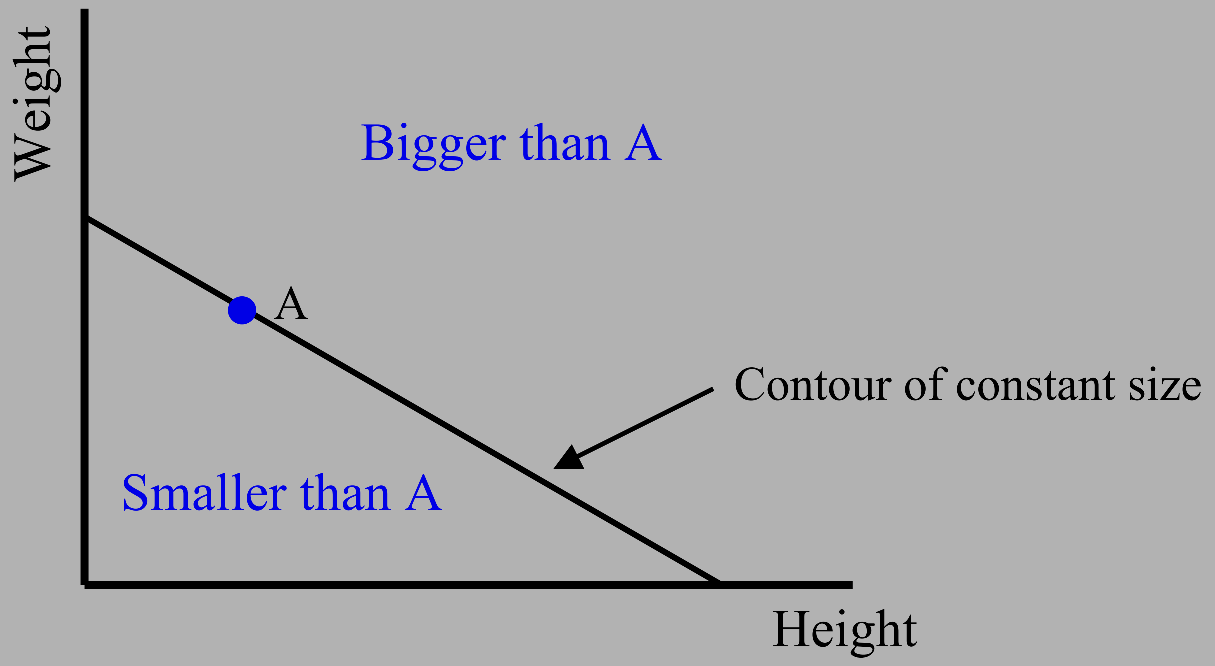
Combining Indicators

Simple Average

$$I_1 = \text{Height}$$

$$I_2 = \text{Weight}$$

$$\text{Size} = (I_1 + I_2)/2$$



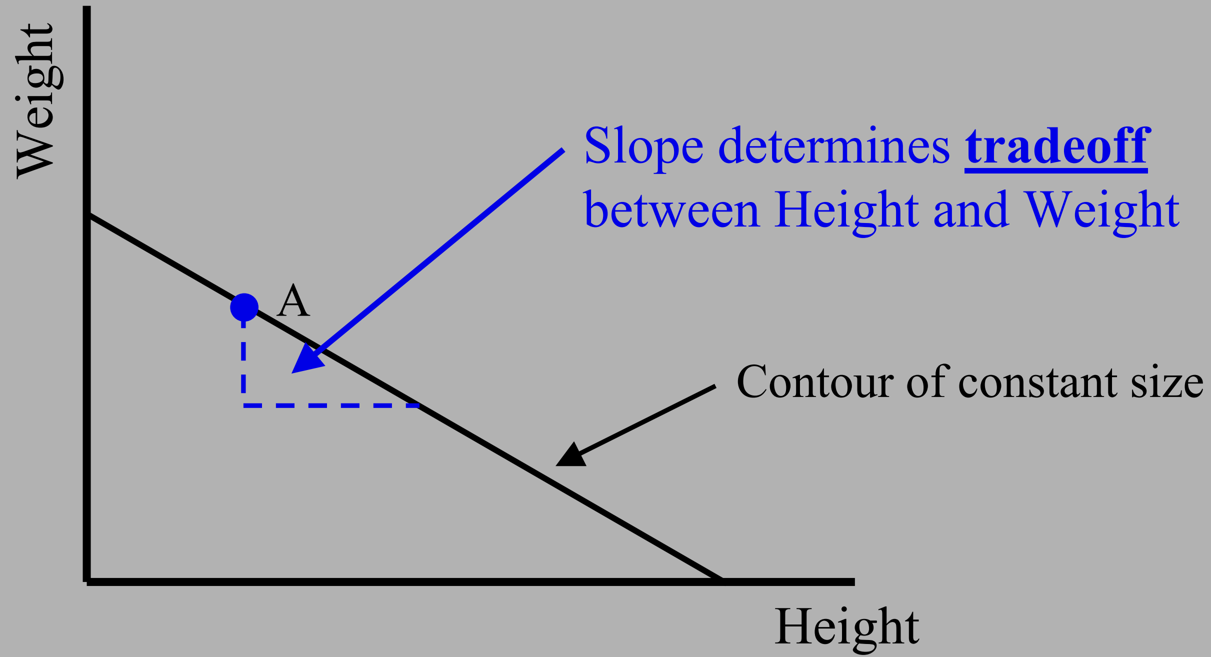
Combining Indicators

Weighted Average

$$I_1 = \text{Height}$$
$$I_2 = \text{Weight}$$
$$\text{Size} = w_1 I_1 + w_2 I_2, \qquad w_1 + w_2 = 1$$

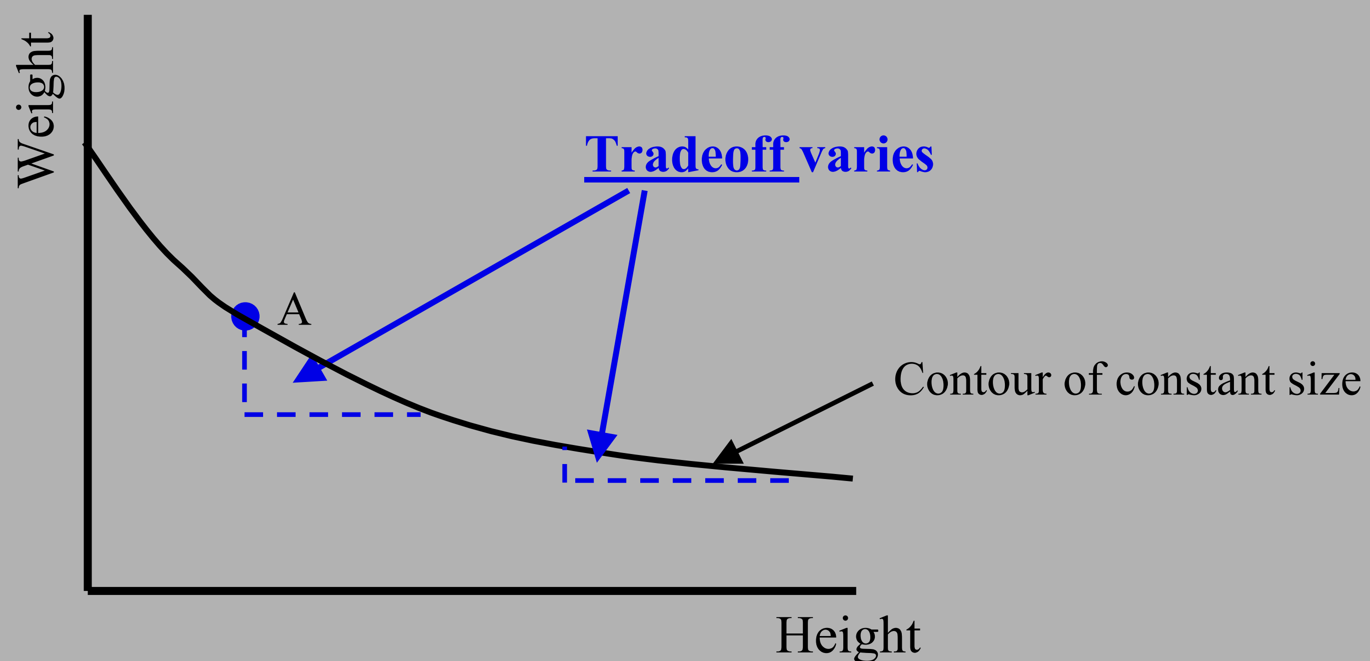
w_1 and w_2 reflect:

- Units of measurement
- Relative importance of the two indicators



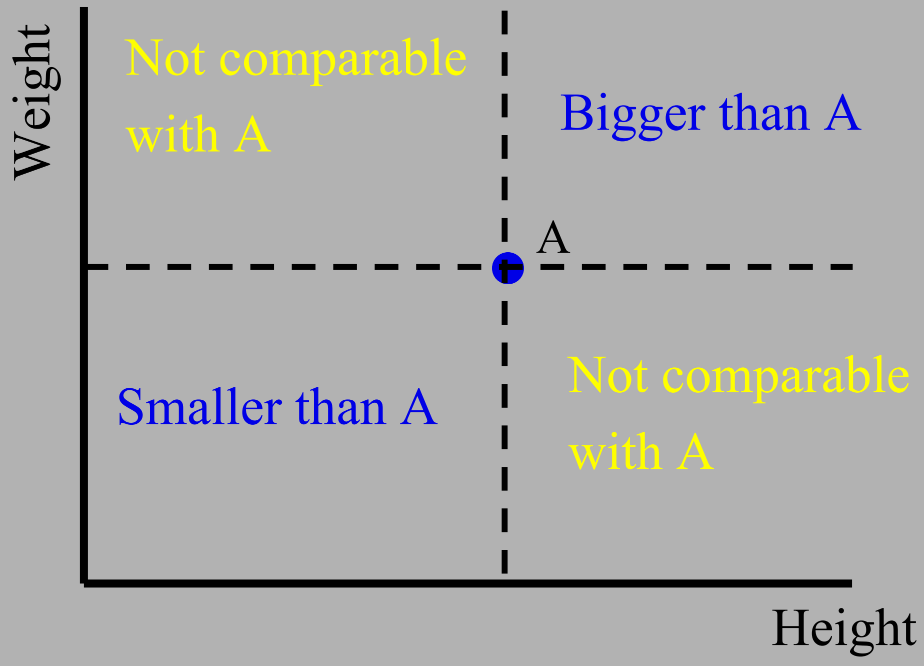
Combining Indicators
Non-Linear Combination

$$I_1 = \text{Height}$$
$$I_2 = \text{Weight}$$
$$\text{Size} = F(I_1, I_2)$$



Partial Ordering

$$I_1 = \text{Height}$$
$$I_2 = \text{Weight}$$



UNEP HEI
National Land, Air, Water Indicators

HEI with revised data:

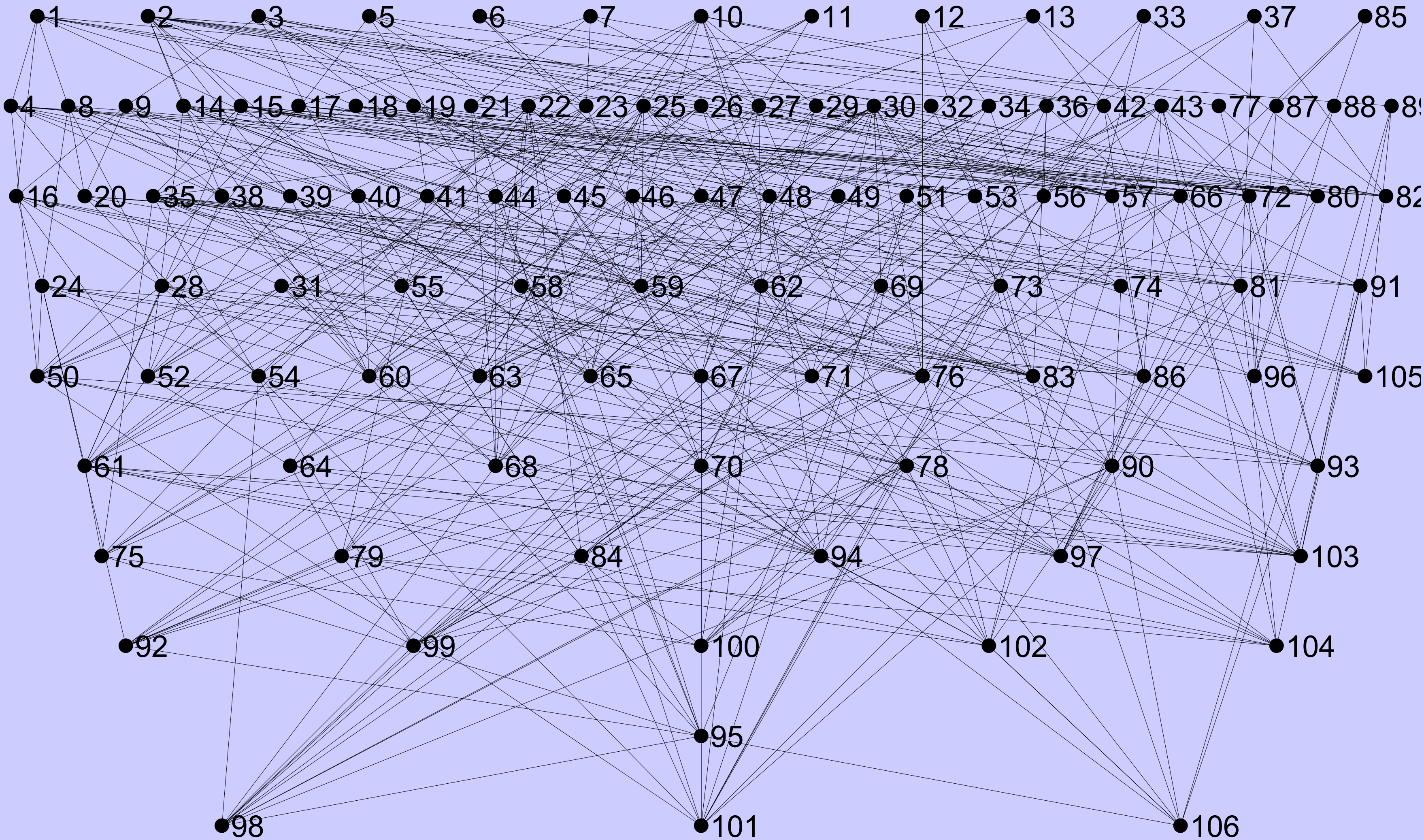
- **Land**: undomesticated land to total land area
- **Air**: (**air indicator1** + **air indicator2**) / 2, where
  air indicator1 = renewable energy use to total energy use;
  air indicator2 = GDP per unit energy use, based on maximum and minimum concept
- **Water**: (**water indicator1** + **water indicator2**) / 2, where
  water indicator1 = ratio of water available after annual withdrawals to internal water resources;
  water indicator2 = ratio of people with access to an improved water source to total population

UNEP HEI Data Matrix

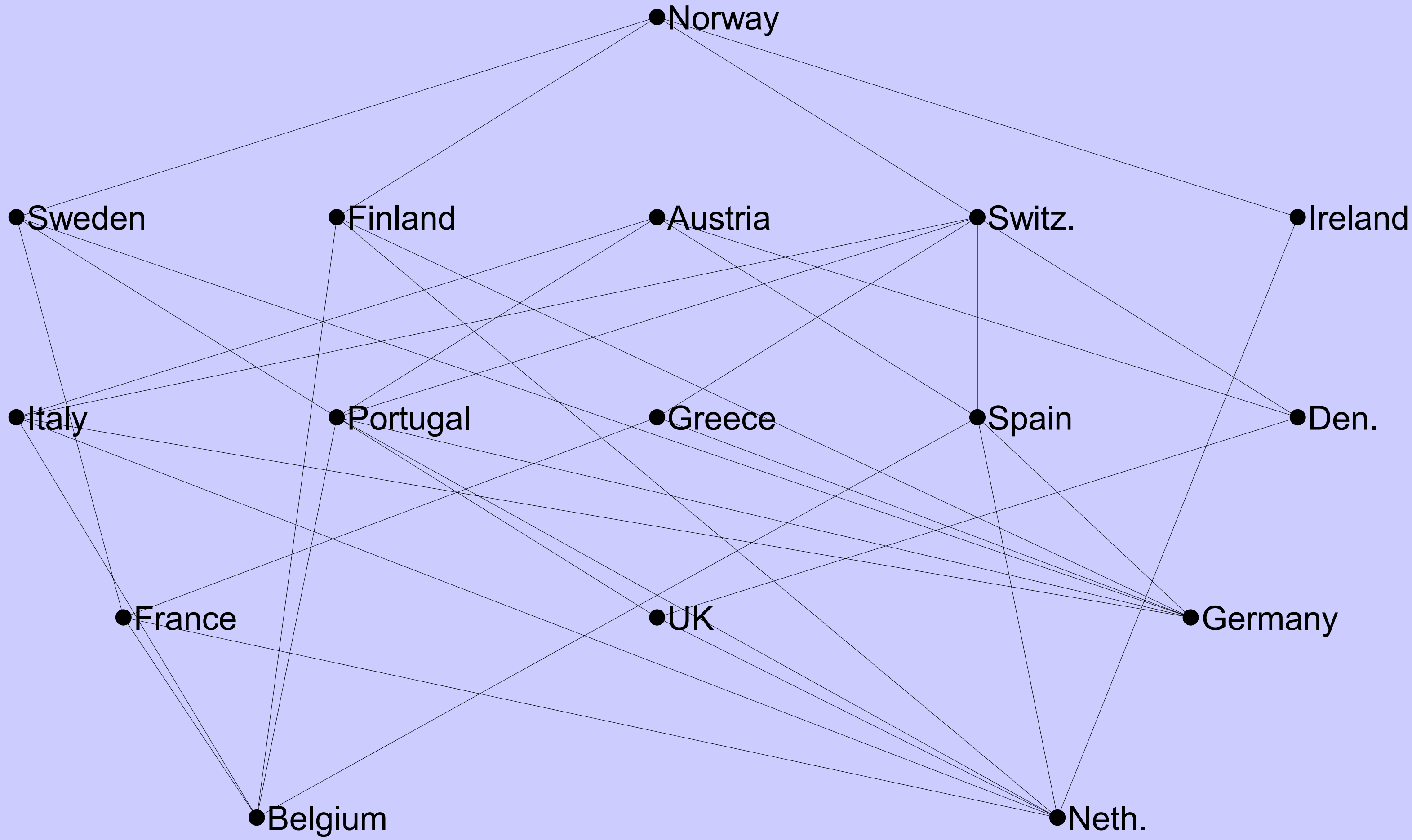
HEI Rank	Country	HEI	Land	Air	Water
1	Costa Rica	0.8503	0.8951	0.6729	0.9829
2	Norway	0.8078	0.9955	0.4305	0.9974
3	Colombia	0.7996	0.9376	0.5087	0.9525
4	Guatemala	0.7988	0.8606	0.5789	0.9569
5	Ghana	0.7856	0.9467	0.5950	0.8150
6	Cameroon	0.7810	0.9665	0.5671	0.8093
7	Peru	0.7806	0.9847	0.5484	0.8088
8	Honduras	0.7764	0.8832	0.5097	0.9363
9	Brazil	0.7759	0.9112	0.4850	0.9315
10	Gabon	0.7757	0.9977	0.4797	0.8498
11	Benin	0.7670	0.9888	0.5042	0.8080
12	Albania	0.7537	0.6411	0.7500	0.8700
13	Congo, Dem. Rep.	0.7528	0.9833	0.5503	0.7248
14	Chile	0.7526	0.9487	0.3571	0.9521
15	Sweden	0.7519	0.9818	0.2824	0.9917

Hasse Diagram---All Countries

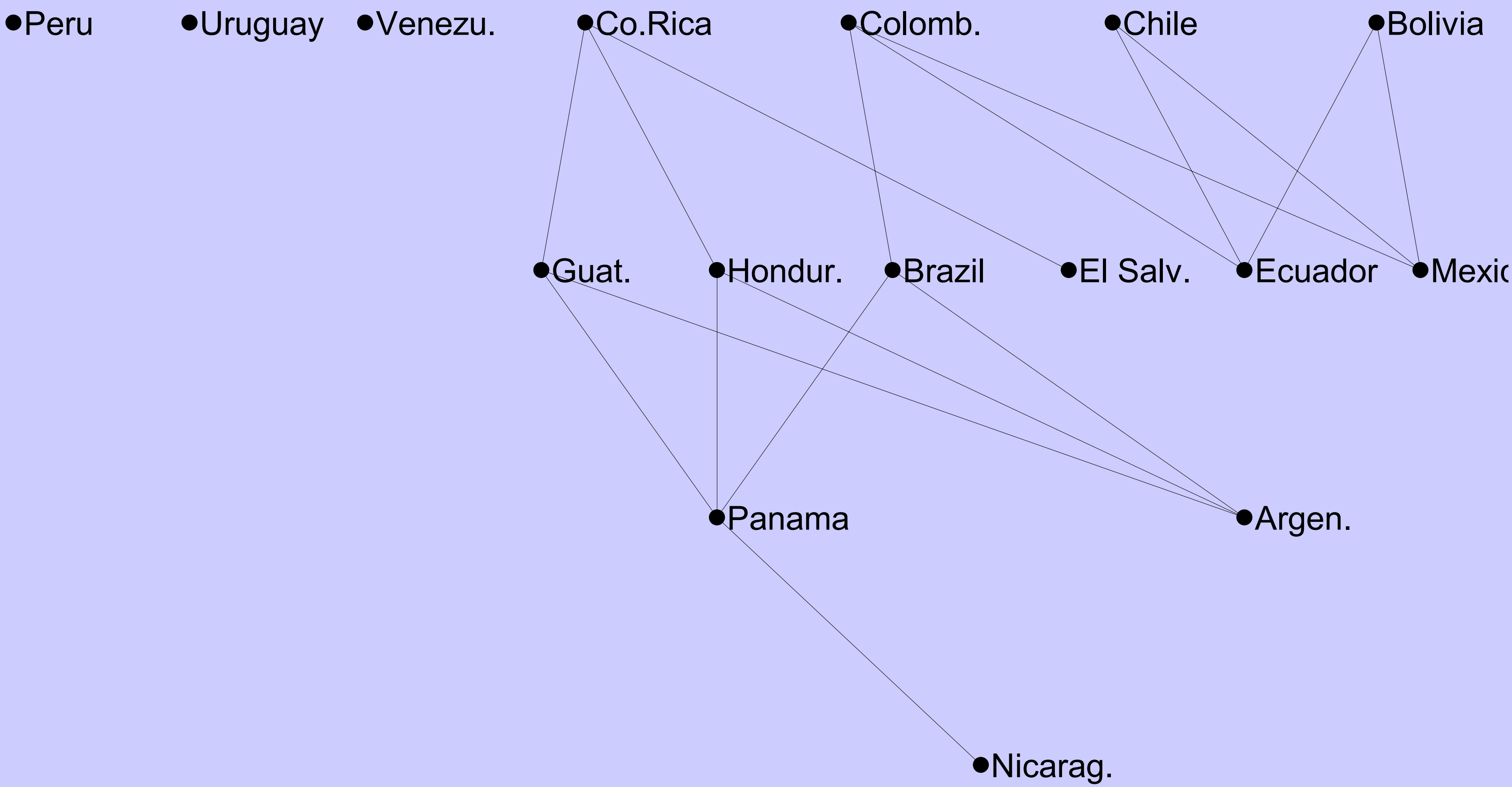
(labels are HEI ranks)



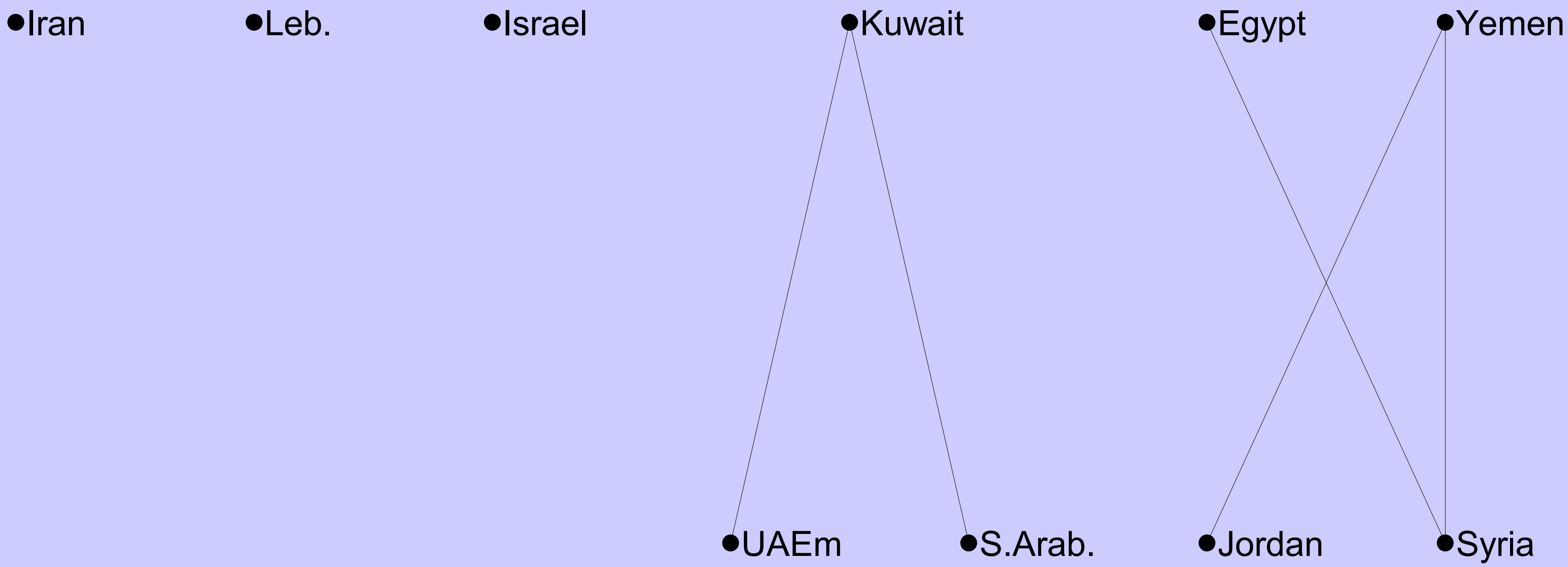
Hasse Diagram --- Western Europe



Hasse Diagram --- Latin America



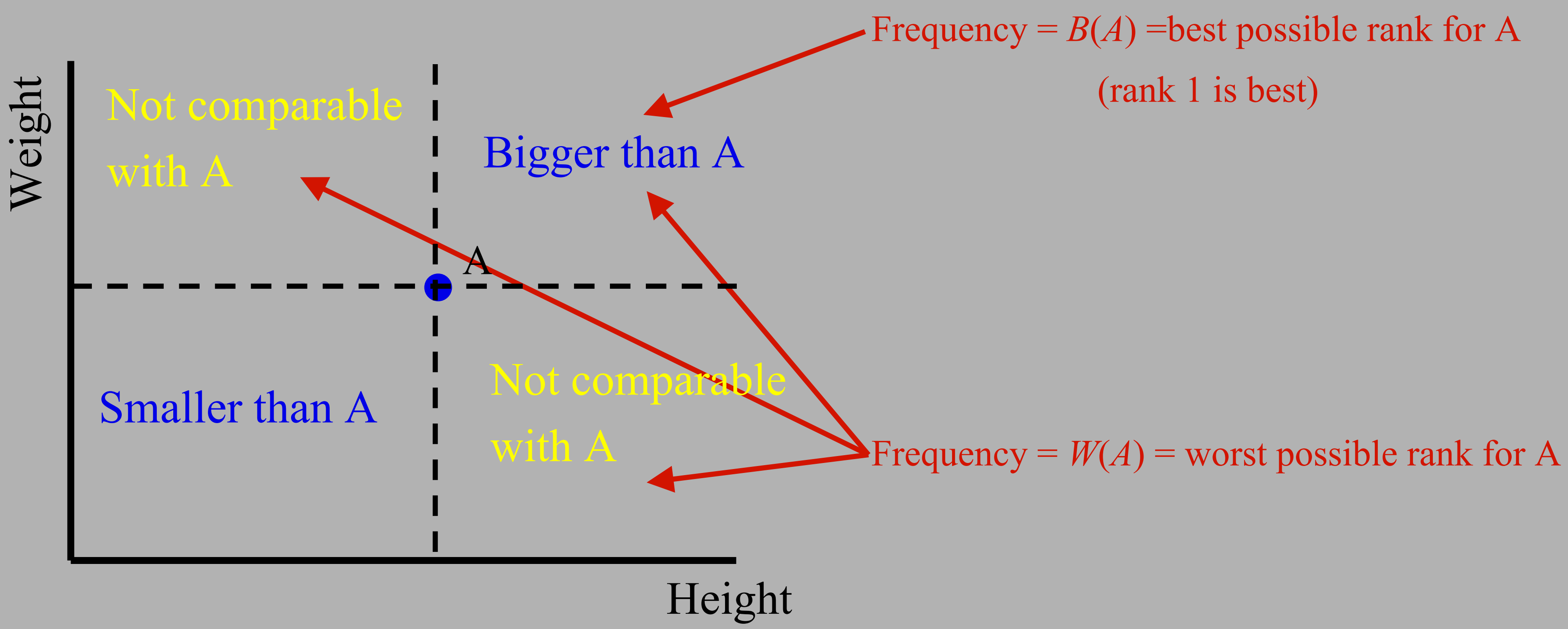
Hasse Diagram --- Middle East



Rank Interval of a Country

- The HEI ranks the countries of the world in a manner that is consistent with the Land, Air, and Water indicators
- There are many other consistent rankings possible
- As we vary over all the consistent rankings, each country receives a set of possible ranks. This set of ranks turns out to be an interval of consecutive integers and is called the **Rank Interval** for the country.
- A wide rank interval indicates much ambiguity or disagreement when the country in question is compared with other countries
- Conversely, a narrow interval indicates general consensus regarding a country's comparative standing

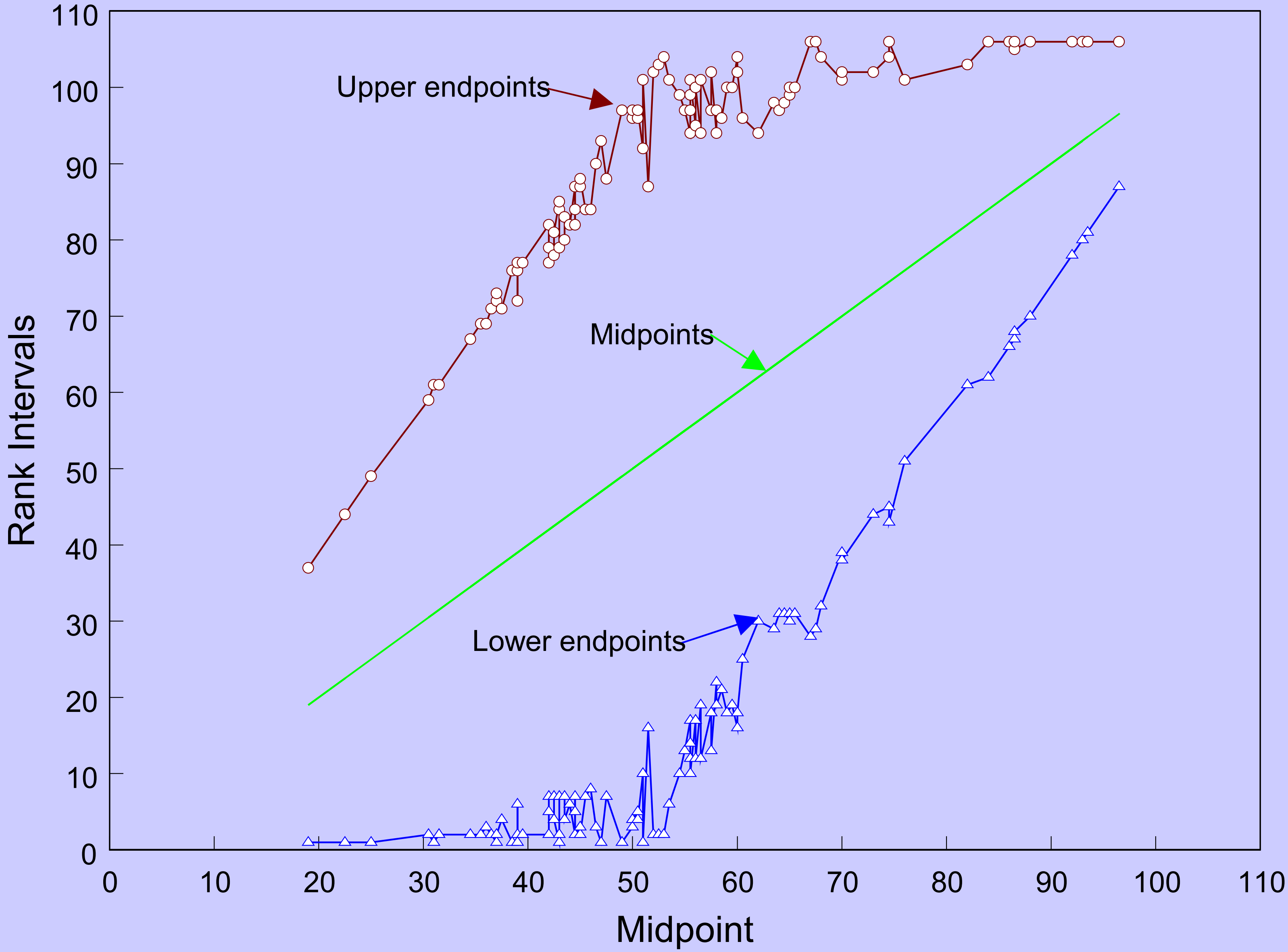
Computing Rank Intervals



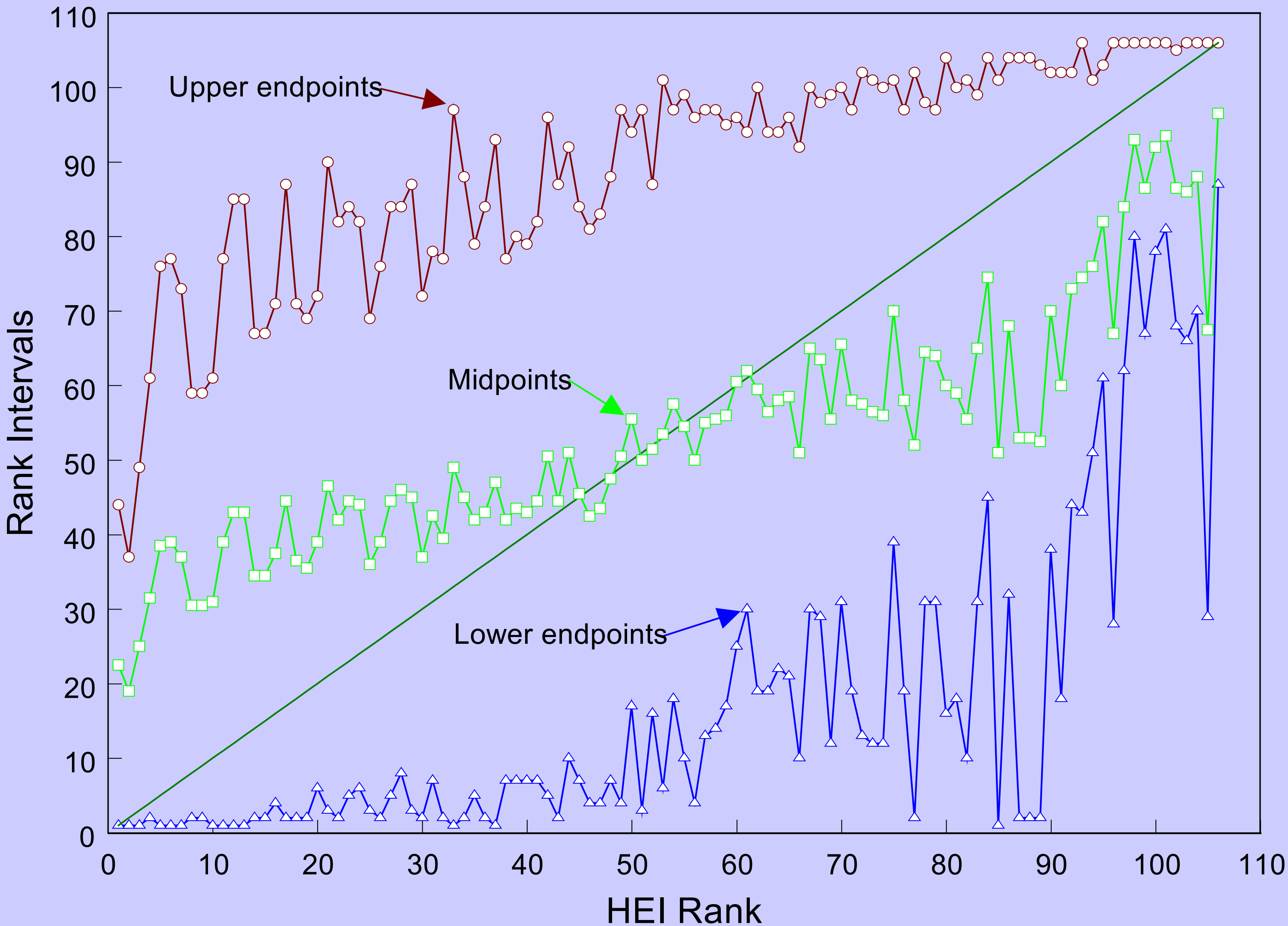
Rank Interval for *A*: all integers r such that $B(A) \leq r \leq W(A)$

Ambiguity for *A*: length of rank interval = $W(A) - B(A)$

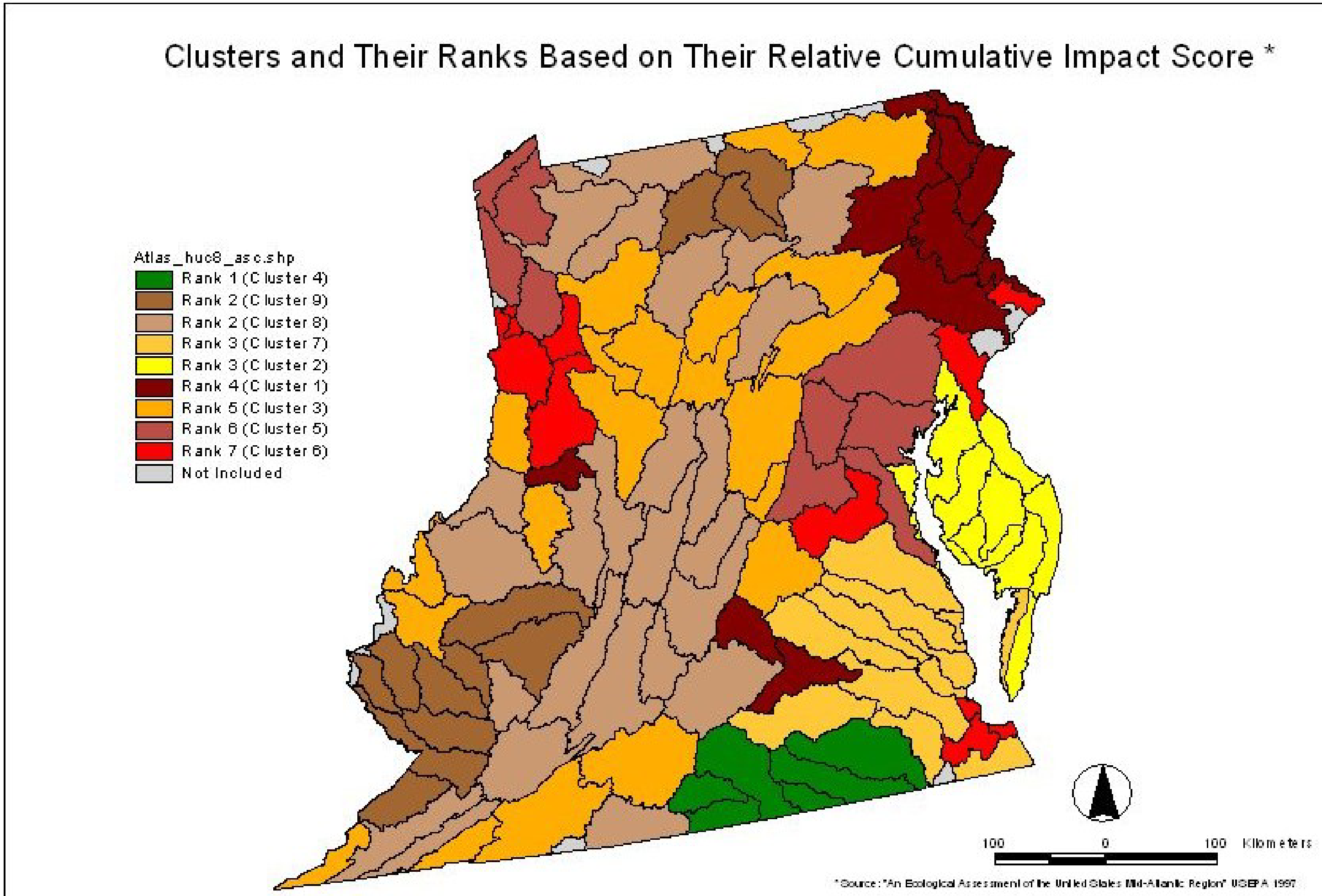
Rank Intervals vs Midpoints



Rank Intervals vs HEI Ranks



Watersheds of Mid-Atlantic Region



Watersheds of Mid-Atlantic Region

- 114 Watersheds (8-digit HUCs)
- 9 indicators used here

- Data Matrix (extract):

HUC	POPDENS	POPCHG	RDDENS	SO4DEP	RIPFOR	STRD	DAMS	CROPSL	INTALL
2040101	51.72	54	2.74	2050	13.84	6.51	32.33	5.99	69.48
2040103	114.53	35	1.7	2087	8.06	6.05	85.41	5.59	77.93
2040104	76.16	81	3.61	2150	7.46	5.69	86.57	1.75	38.78
2040105	625.06	25	7.71	2443	22.94	9.68	37.64	13.68	91.49
2040106	488.13	13	2.73	2435	16.76	7.18	42.02	7.5	72.3
2040201	1291.56	20	6.92	2634	28.83	6.87	47.12	5.48	100
2040202	3533.76	-5	16.01	2817	33.71	10.65	49.04	2.11	99.93
2040203	816.35	-1	3.32	2799	30.94	10.15	45.89	9.46	93.23
2040205	596.15	15	3.17	2913	26.65	6.68	12.17	5.4	100

Watersheds of Mid-Atlantic Region

Indicator Description

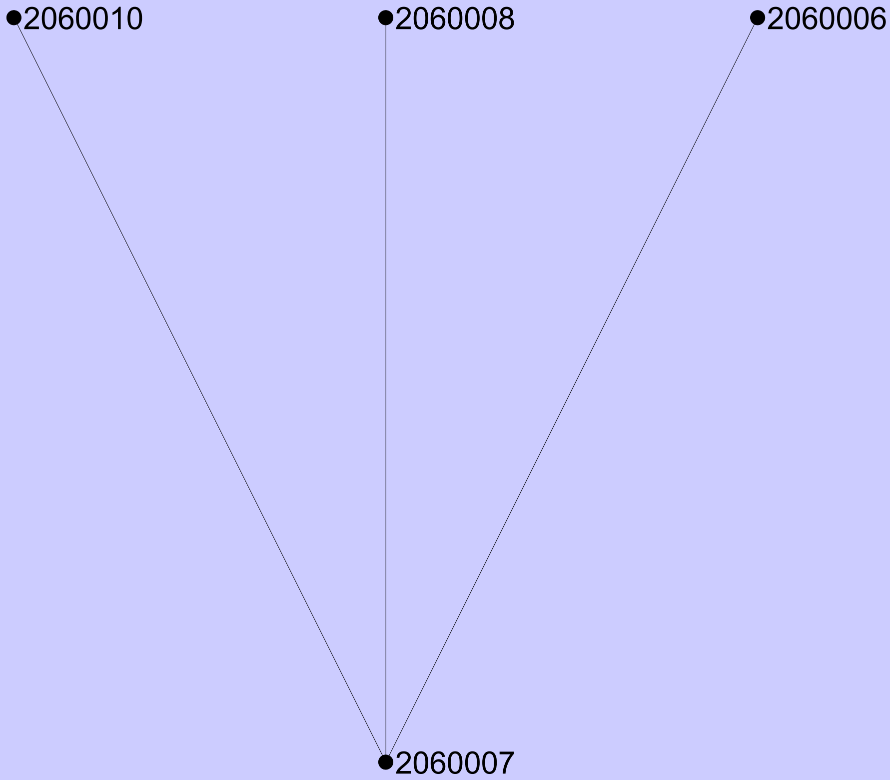
POPDENS	1990 population density
POPCHG	1970-1990 population change
RDDENS	Road density
SO4DEP	Sulfate deposition
RIPFOR	Proportion forested stream-length
STRD	Proportion stream-length with nearby roads
DAMS	Impoundments per 1000 km stream-length
CROPSL	Proportion of WS with crops on >3 percent slope
INTALL	Proportion of WS with interior forest habitat

Watersheds of Mid-Atlantic Region
Hasse Diagrams

60 connected components in Hasse Diagram:

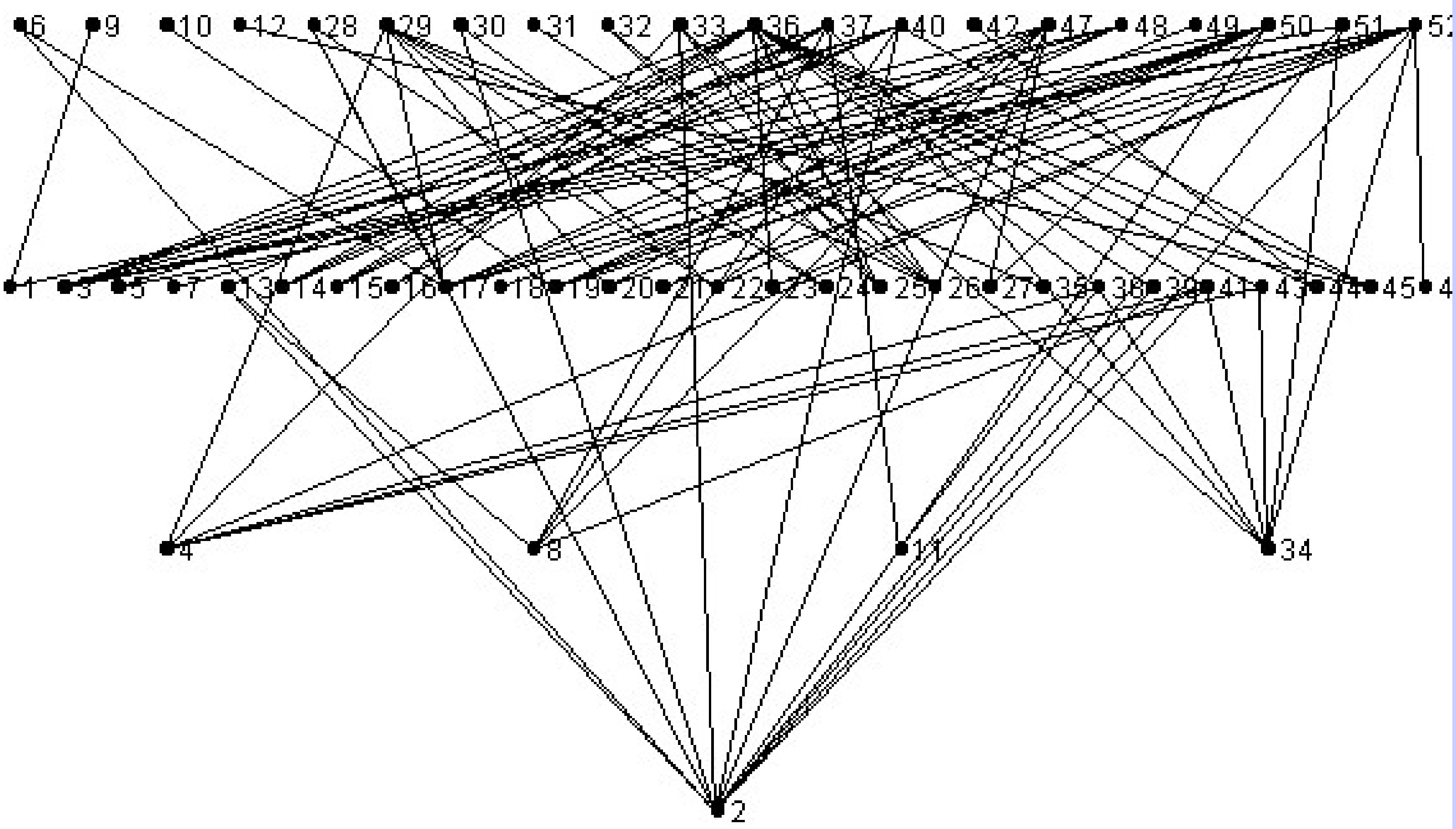
- 58 of the components are isolates
- 1 component (**secondary component**) contains 4 watersheds
- 1 component (**primary component**) contains 52 watersheds

Watersheds of Mid-Atlantic Region
Hasse Diagram for Secondary Component



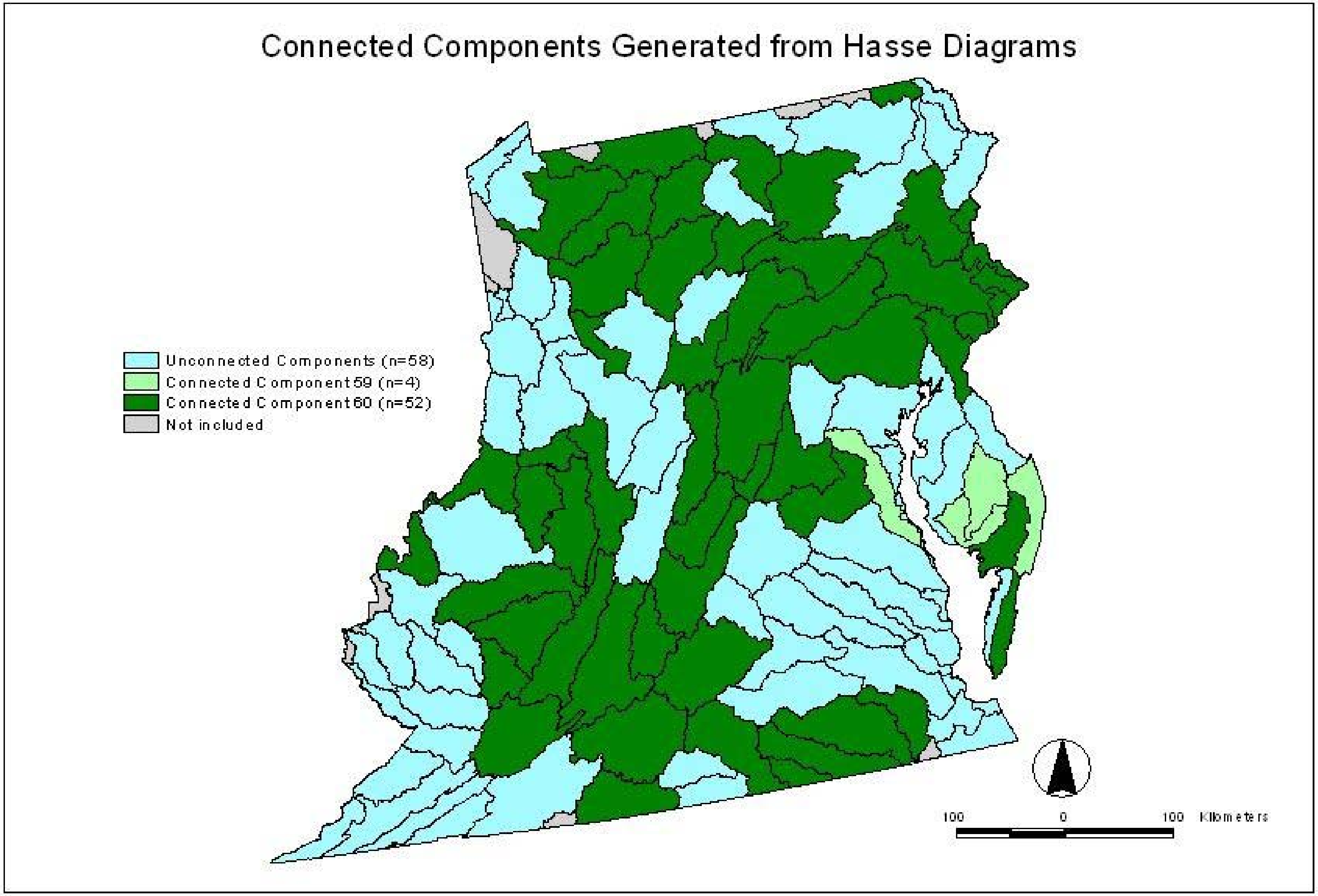
Note: Labels are HUC numbers

Watersheds of Mid-Atlantic Region
Hasse Diagram for Primary Component



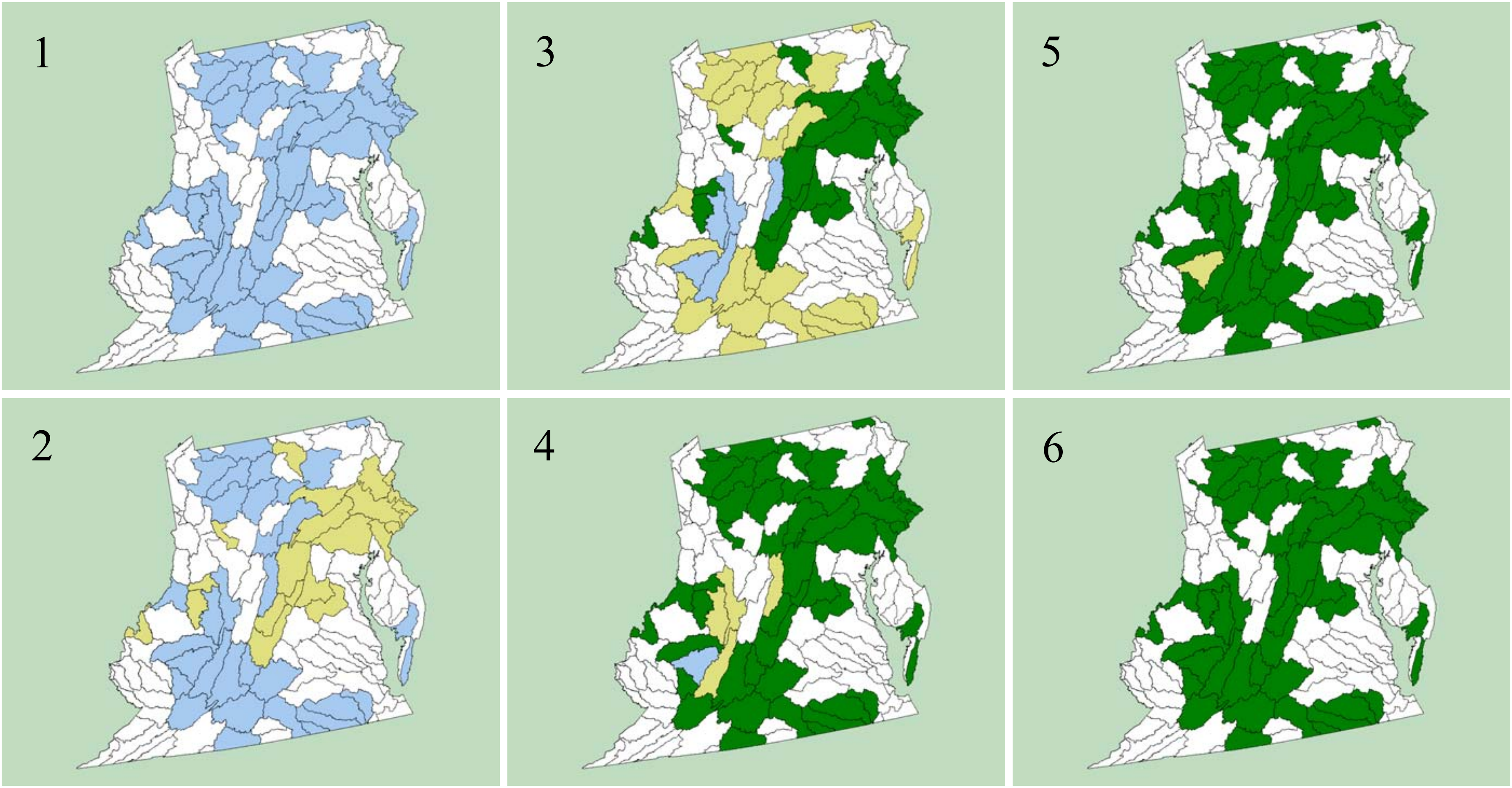
Note: Labels are row numbers in data matrix

Watersheds of Mid-Atlantic Region



Watersheds of Mid-Atlantic Region

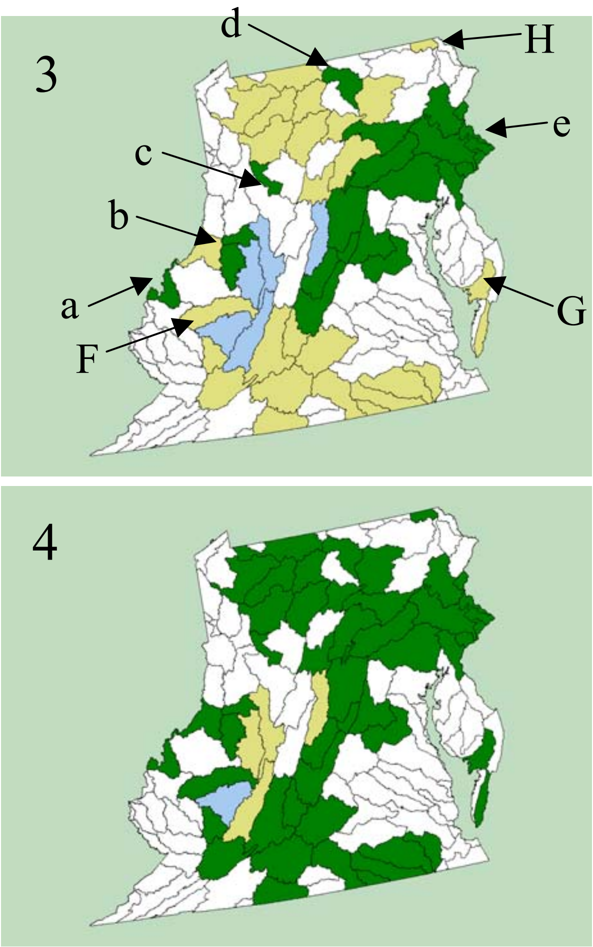
Echelon Analysis of Primary Hasse Connected Component

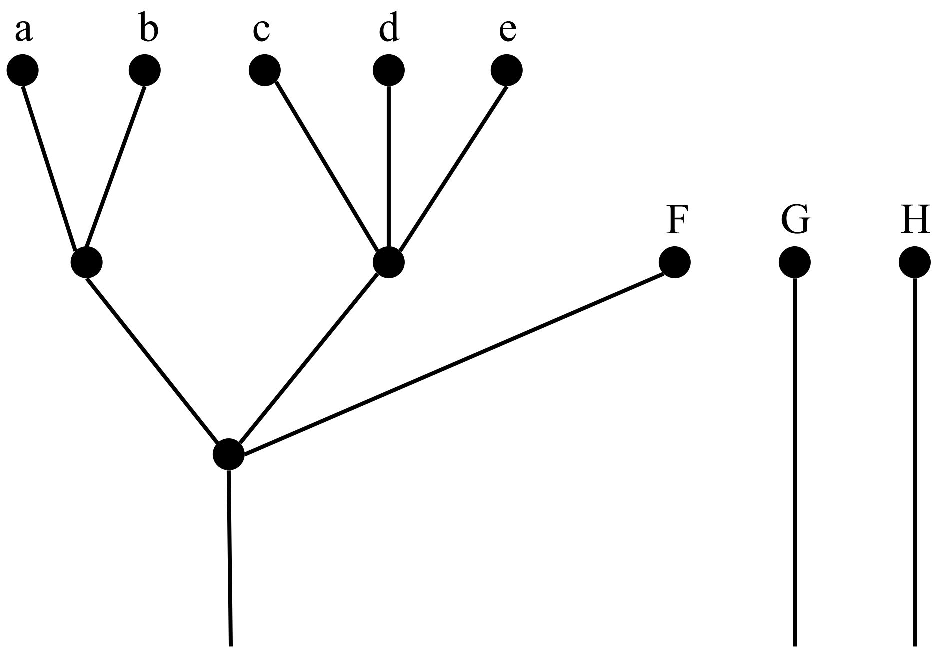


Blue: Underwater **Sandy:** Newly exposed (beach) **Green:** Previously exposed

Watersheds of Mid-Atlantic Region

Echelon "Tree" for Primary Hasse Connected Component





Note: Echelon "Tree" is really a "Forest" because the primary Hasse connected component is geographically disconnected (with three pieces)

Watersheds of Mid-Atlantic Region

Rank-Intervals for Primary Hasse Connected Component

To be prepared

Ranking Partially Ordered Sets – 1

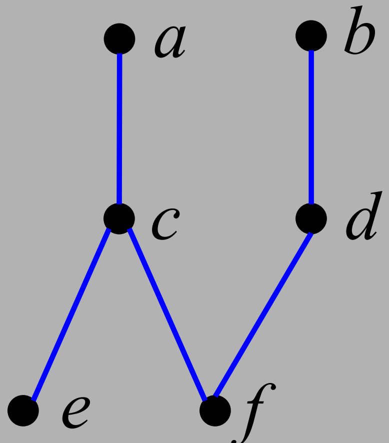
- S = partially ordered set (**poset**) with elements $a, b, c, \ldots.$
- How can we rank the elements of S consistent with the partial order? Such rankings are called **linear extensions** of the partial order.
- Different people with different perceptions and priorities may choose different rankings.
- How many rankings assign rank 1 to element a ? Rank 2 ? Rank 3 ? , etc.
- If rankings are chosen randomly (equal probability), what is the likelihood that element x receives rank i ?

Ranking Partially Ordered Sets – 2

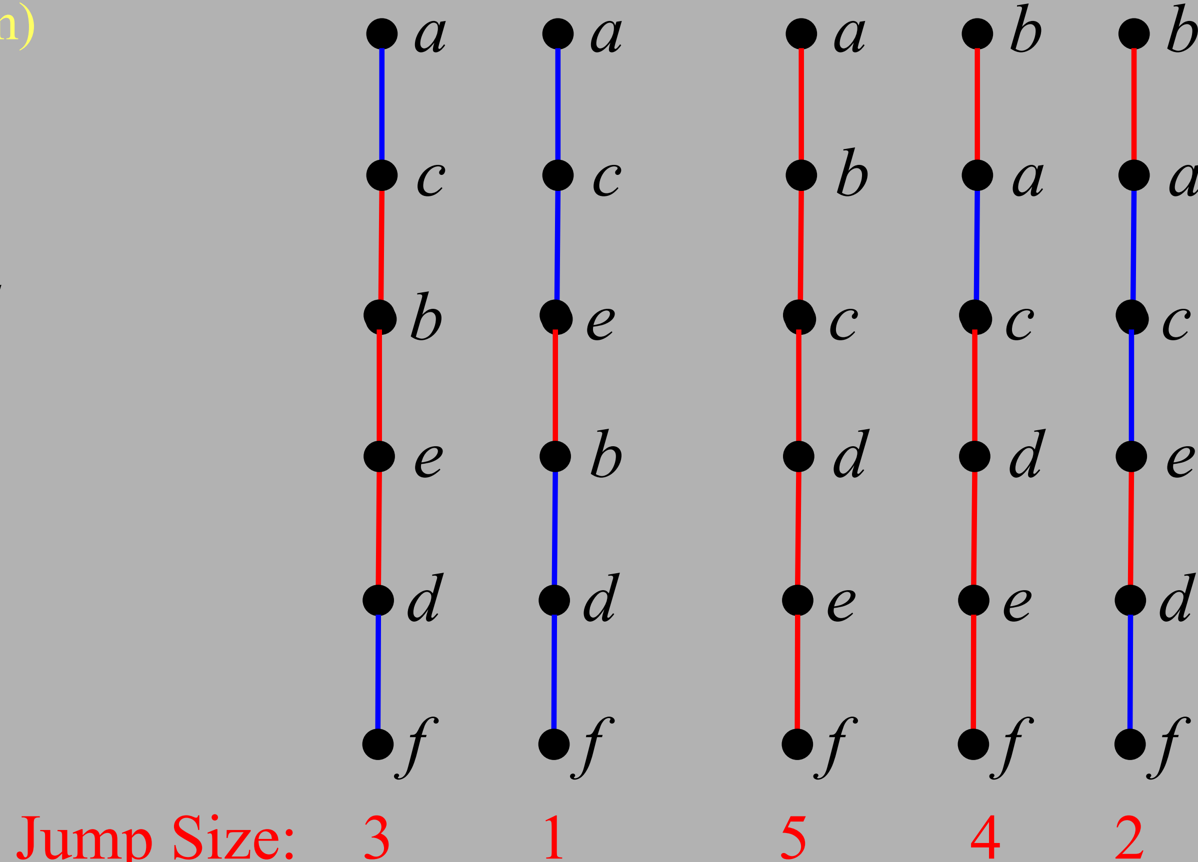
An Example

Poset

(Hasse Diagram)



Some linear extensions



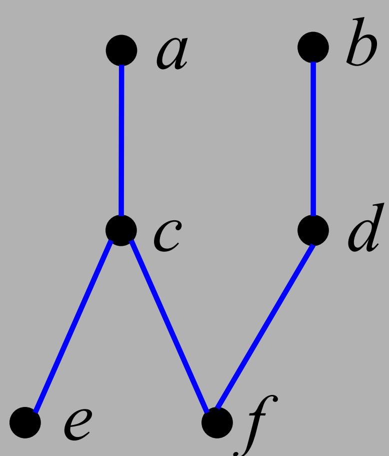
Jump or **Imputed Link** (-------) is a link in the ranking that is not implied by the partial order

Ranking Partially Ordered Sets – 2a

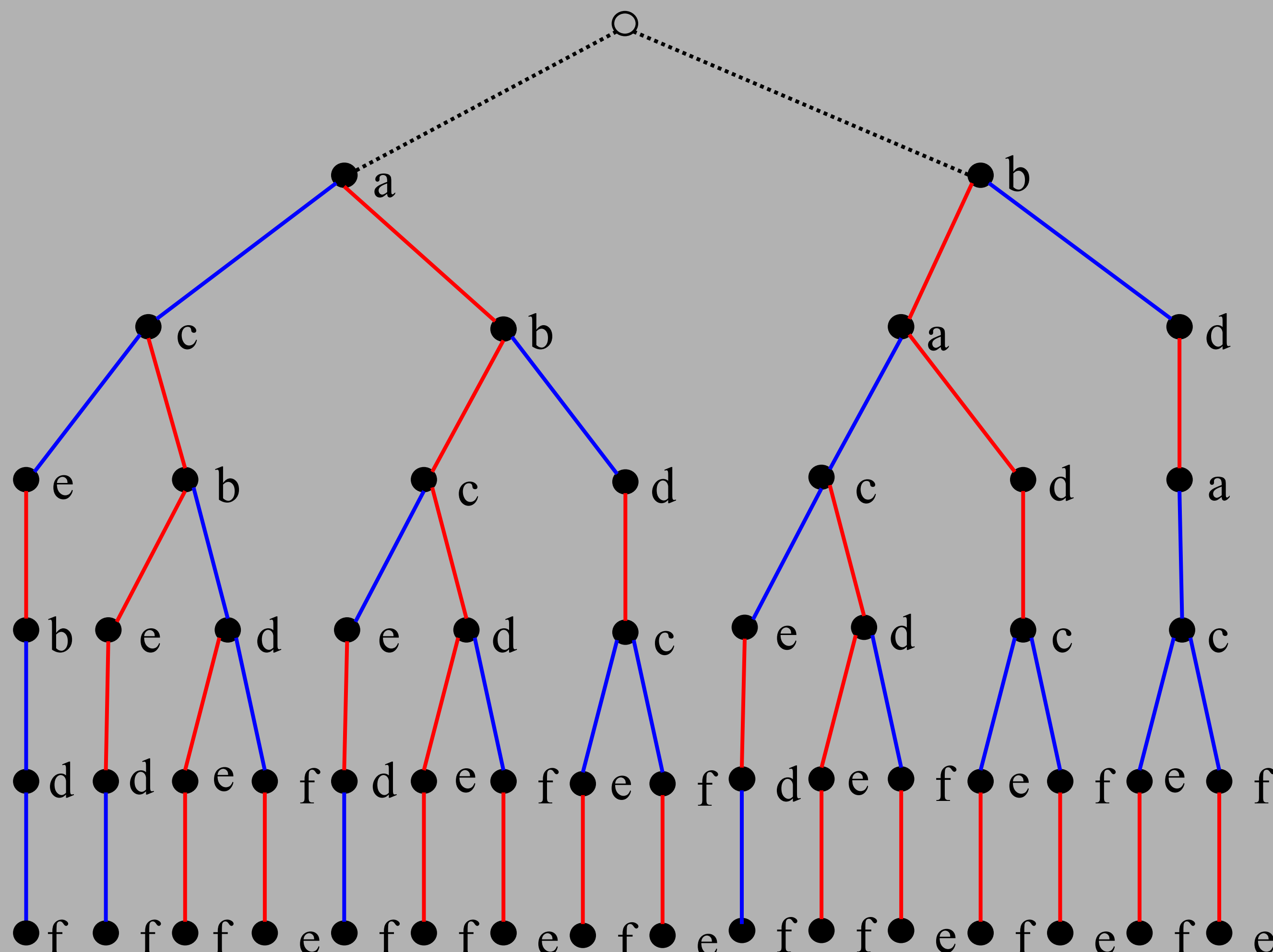
- How can we enumerate all possible linear extensions of a given poset (16 for the example) ?
- All possibilities can be laid out in a tree diagram, called the **linear extension decision tree**:
  - Start from the root node
  - Select any **maximal** element from the poset
  - Remove that element from the poset
  - Select any **maximal** element from the reduced poset
  - Continue selecting and removing maximal elements until the poset is exhausted
- There is a one-to-one correspondence between paths through the tree and the linear extensions

Ranking Partially Ordered Sets – 2b

Poset
(Hasse Diagram)



Linear extension decision tree



Jump Size: 1 3 3 2 3 5 4 3 3 2 4 3 4 4 2 2

Ranking Partially Ordered Sets – 3

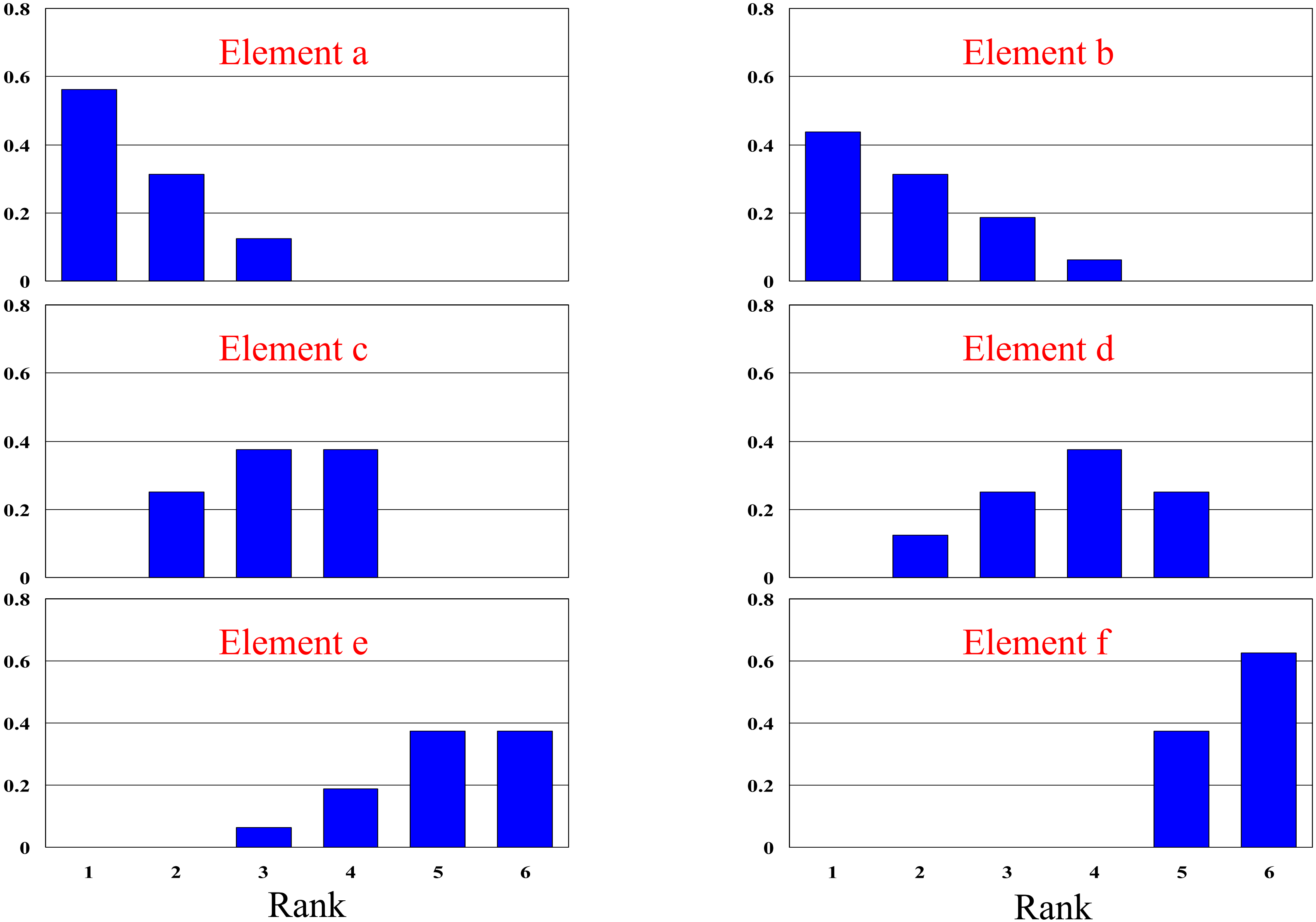
In the example from the preceding slide, there are a total of 16 linear extensions, giving the following frequency table.

	Rank						
Element	1	2	3	4	5	6	Totals
a	9	5	2	0	0	0	16
b	7	5	3	1	0	0	16
c	0	4	6	6	0	0	16
d	0	2	4	6	4	0	16
e	0	0	1	3	6	6	16
f	0	0	0	0	6	10	16
Totals	16	16	16	16	16	16	

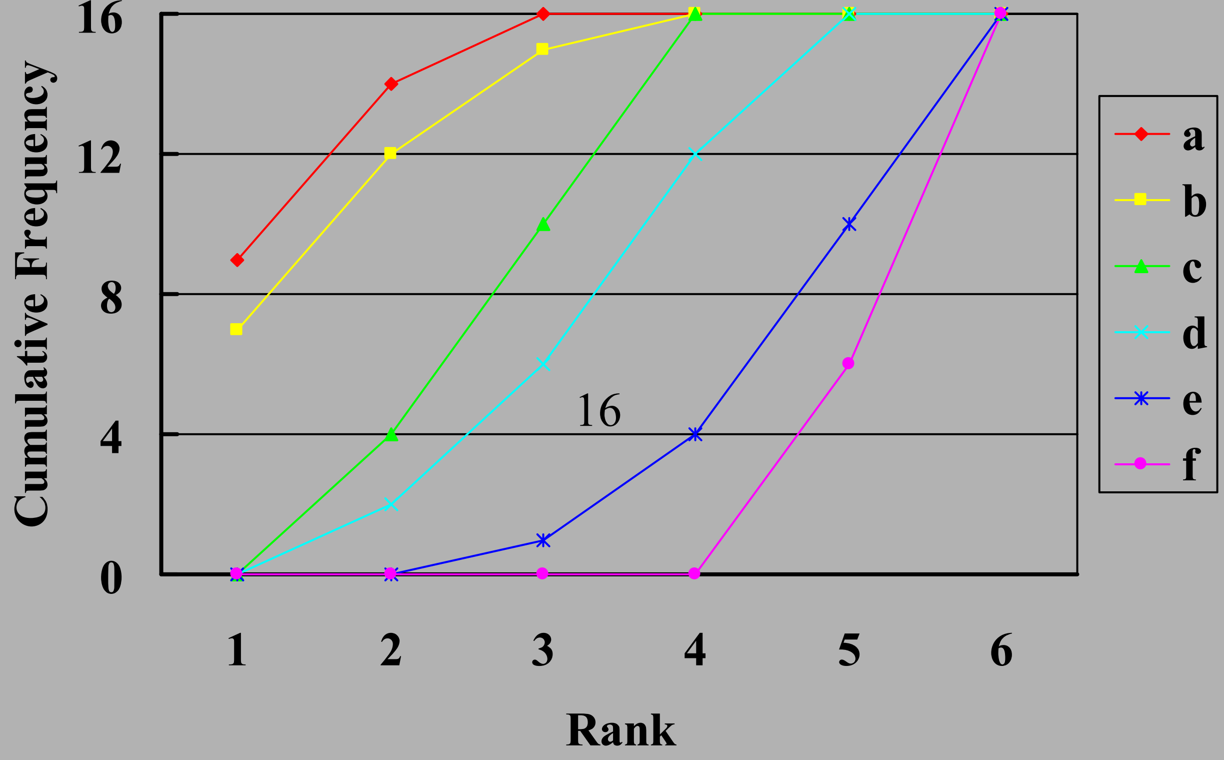
- Each (normalized) row gives the rank-frequency distribution for that element
- Each (normalized) column gives a rank-assignment distribution across the poset

Ranking Partially Ordered Sets – 3a

Rank-Frequency Distributions



Ranking Partially Ordered Sets – 3b



The curves are stacked one above the other and the result is a linear ordering of the elements: $a > b > c > d > e > f$

Ranking Partially Ordered Sets – 3c

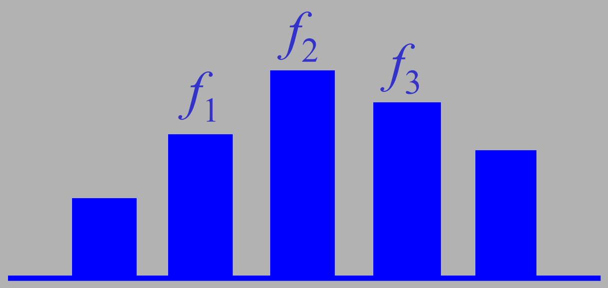
Properties of Rank-Frequency Distributions

- The rank-frequency distributions for any poset are unimodal
- In fact, there is a theorem which asserts that each rank-frequency distribution is log-concave, i.e., if

  $$f_1 ,\ f_2 ,\ \text{and}\ f_3$$

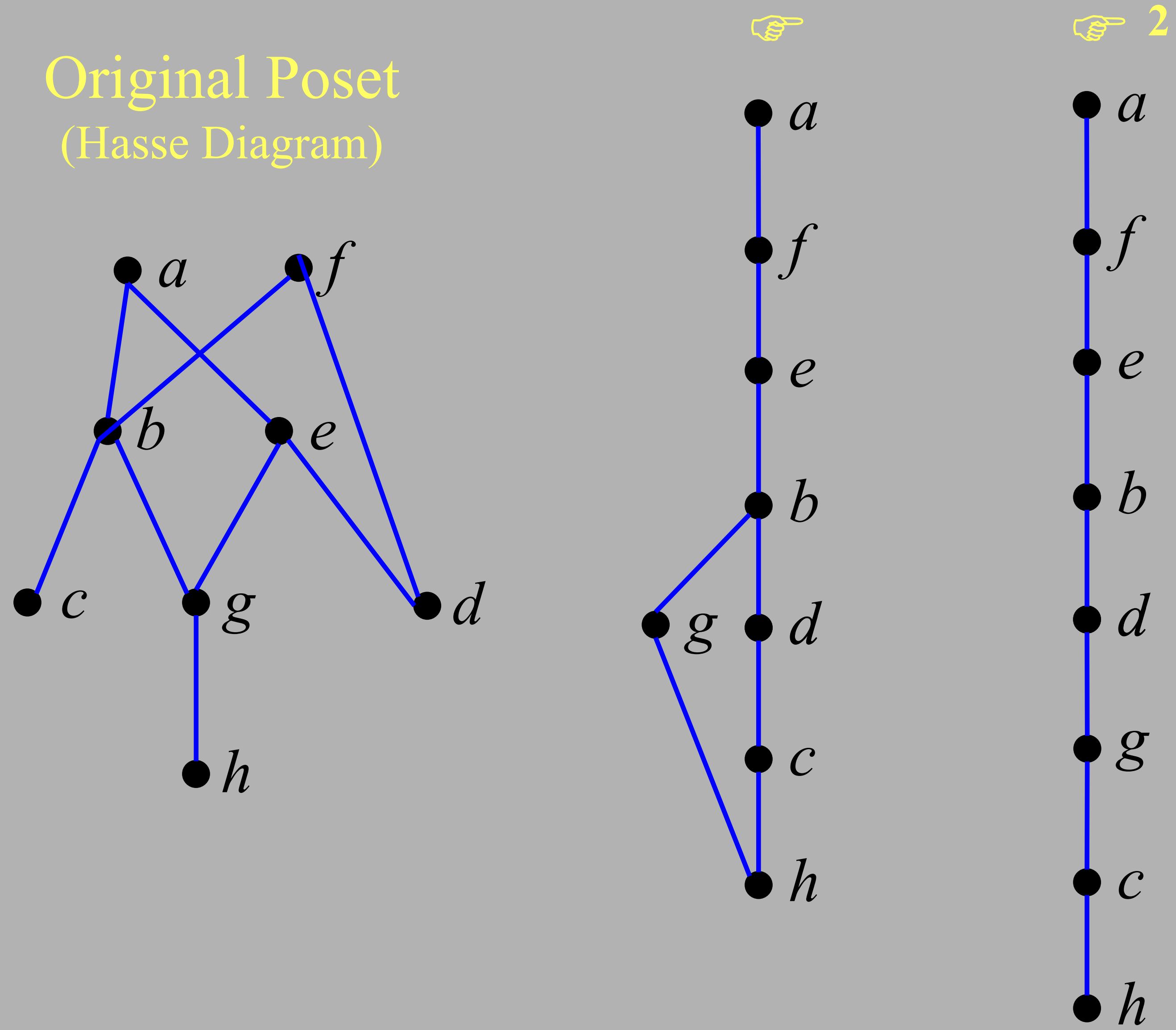
  are the frequencies (relative or absolute) for any three consecutive ranks assigned to an element of the poset, then

  $$(f_2)^2 \geq f_1 \cdot f_3$$



Cumulative Rank Frequency Operator –1

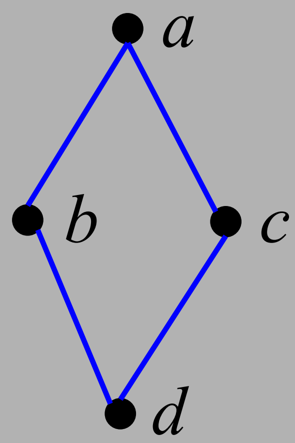
An example where ☞ must be iterated



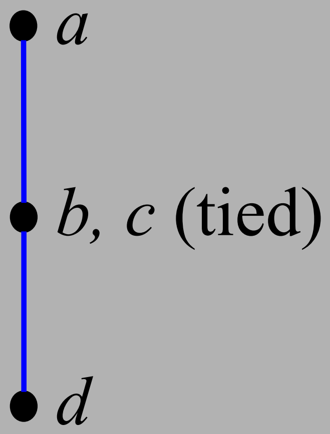
Cumulative Rank Frequency Operator – 2

An example where ☞ results in ties

Original Poset
(Hasse Diagram)



☞



- Ties reflect symmetries among incomparable elements in the original Hasse diagram

- Elements that are comparable in the original Hasse diagram will not become tied after applying ☞ operator

Cumulative Rank Frequency Operator – 3

- In most cases of practical interest, the number of linear extensions, $e(S)$, is too large for actual enumeration in a reasonable length of time.
- For the HEI data set, the number of linear extensions satisfies:

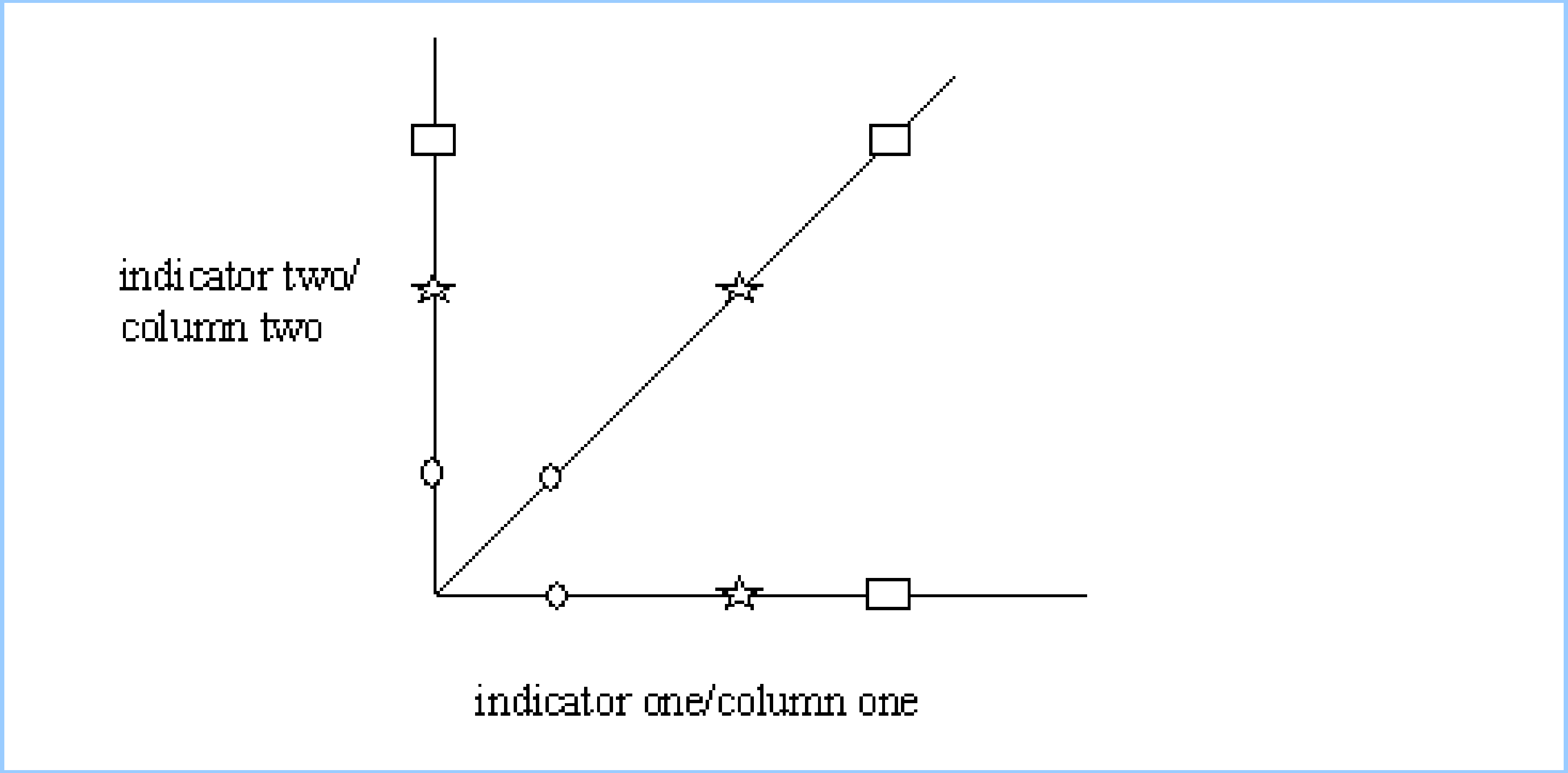
$$8.6 \times 10^{105} \leq e(S) \leq 1.9 \times 10^{243}$$

  which is completely beyond present-day computational capabilities.
- So what do we do?
- **Markov Chain Monte Carlo (MCMC)** applied to the uniform distribution on the set of all linear extensions lets us **estimate** the normalized rank-frequency distributions. Estimating the absolute frequencies (approximate counting) is also possible but somewhat more difficult.

More Indicators do not necessarily mean **More Information** and **More Discrimination**

- Suppose two indicator columns are exactly the same.
- The second column does not add new information or discriminatory capability to that of the first column.
- Likewise, there is little new information when there is a strong underlying (rank-)correlation between the two indicators.
- Landscape ecologists have discovered that some fifty landscape pattern metrics amount to essentially five to ten indicators.

Multiple Indicators can convey **Redundant Information**



Combining Indicators (e.g., by averaging) typically **fails** to adjust for the **Redundancy**

Logo for Statistics, Ecology, Environment, and Society