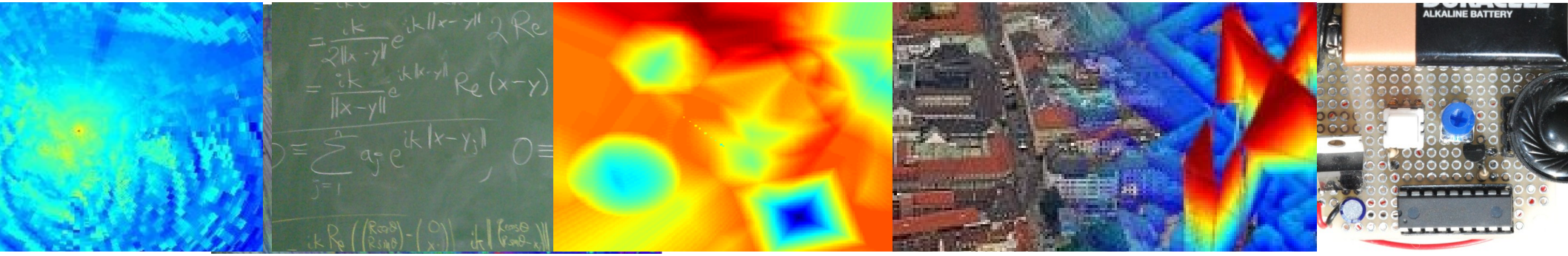
The Category of Binary Relations, Dowker complexes, Cosheaves, and Functoriality



Michael Robinson

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001119C0072. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of DARPA.

Acknowledgments

- Collaborators:
  - Kris Ambrose, Steve Huntsman, Allyson O'Brien, Matvey Yutin
- Students:
  - Ken Ewing (MS Math Info & Sec)
- Funding:
  - Sergey Bratus (DARPA/I2O) SafeDocs

- Key reference:

  M. Robinson, "Cosheaf representations of relations and Dowker complexes."

  https://arxiv.org/abs/2005.12348

  (basically everything in these slides comes from this preprint)

Motivation: Consensus file formats

- What does it mean for files to comply with a format, especially if the format is an ambiguous, community-defined consensus?
- <u>Tactic</u>: Use a *binary relation,* recording which files are accepted as valid by which parsers
- <u>Hypothesis</u>: Anomalous files or parsers will manifest within the context of this relation

$$
\begin{array}{cc}
 & \text{files} \\
\text{parsers} & \begin{array}{c} A \\ B \\ C \\ D \end{array}
\begin{pmatrix}
10000011111100001111 \\
01100011100010000000 \\
00011000010011111111 \\
00000100001101111111
\end{pmatrix}
\end{array}
$$

1 = file parsed successfully
0 = problem parsing file

It's probably the case that there are many more files than parsers, but this isn't terribly crucial

Main ideas of the talk

- The category **Rel** of relations has a simplicial representation, the *Dowker complex*
  - The Dowker complex is a covariant functor
  - **Rel** applies whenever you have tabular data
  - Files accepted by parsers, extant key-value pairs, etc.
- The Dowker complex isn't a complete invariant, but it is when *weighted*
  - May enable statistical analysis of the topology within and between tables
- This leads to a faithful *cosheaf representation*
  - Topology is therefore a reliable description of a table
- The cosheaf carries both the Dowker complex and its *dual* (transpose of the relation)
  - One of the two Dowker complexes will usually be **much** bigger
  - You can use the cosheaf on the smaller complex without loss of structure

Binary relations in their category

- We should formalize the category **Rel** of relations
- Objects are triples (X,Y,R) where $R \subseteq X \times Y$
- These are best thought of as tables of booleans*

$$\begin{array}{c c} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} & \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

$X = \{a, b, c, d, e\}$, $Y = \{1, 2, 3, 4, 5\}$

$R = \{(a,1), (a,2), (b,1), (b,3), (c,2), (c,3), (c,4), (c,5), (d,3), (d,4), (e,4), (e,5)\}$

*This generalizes neatly to entries with poset values

Morphisms in **Rel**

- A morphism in Rel consists of a pair of functions

  $\boldsymbol{f : X_1 \to X_2}$, $g : Y_1 \to Y_2$

  satisfying a compatibility condition

$$
\begin{array}{c}
a \\ b \\ c \\ d \\ e
\end{array}
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 1
\end{pmatrix}
\xrightarrow{\;f\;}
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1
\end{pmatrix}
\begin{array}{c}
A \\ B \\ C
\end{array}
$$

Morphisms in **Rel**

- A morphism in Rel consists of a pair of functions

  $f : X_1 \to X_2$, $\boldsymbol{g : Y_1 \to Y_2}$

  satisfying a compatibility condition

$$\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array} \xrightarrow{\;g\;} \begin{array}{c} \begin{array}{ccccc} g(1) & g(2) & g(3) & g(4) & g(5) \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{array}{c} A \\ B \\ C \end{array} \end{array}$$

Morphisms in **Rel**

- A morphism in Rel consists of a pair of functions

  $f : X_1 \to X_2,\ g : Y_1 \to Y_2$

  satisfying a **compatibility condition**

$$
\begin{array}{c}
 \\ a \\ b \\ c \\ d \\ e
\end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
0 & \boxed{1 \quad 1 \quad 1 \quad 1} \\
0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 1
\end{pmatrix}
\end{array}
\xrightarrow{g,\ f}
\begin{array}{c}
\begin{array}{ccccc} g(1) & g(2) & g(3) & g(4) & g(5) \end{array} \\
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
0 & \boxed{\boxed{1 \quad 1 \quad 1 \quad 1}}
\end{pmatrix}
\end{array}
\begin{array}{c}
 \\ A \\ B \\ C
\end{array}
$$

Morphisms in **Rel**

- A morphism in Rel consists of a pair of functions

  $f : X_1 \to X_2,\ g : Y_1 \to Y_2$

  satisfying a **compatibility condition**

$$
\begin{array}{c}
 \\ a \\ b \\ c \\ d \\ e
\end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 \\
0 & 0 & \boxed{1} & \boxed{1} & 0 \\
0 & 0 & 0 & 1 & 1
\end{pmatrix}
\end{array}
\xrightarrow{g,\ f}
\begin{array}{c}
\begin{array}{ccccc} g(1) & g(2) & g(3) & g(4) & g(5) \end{array} \\
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
0 & \boxed{1} & \boxed{1} & \boxed{1} & \boxed{1}
\end{pmatrix}
\end{array}
\begin{array}{c}
 \\ A \\ B \\ C
\end{array}
$$

Morphisms in **Rel**

- A morphism in Rel consists of a pair of functions

  $f : X_1 \to X_2,\ g : Y_1 \to Y_2$

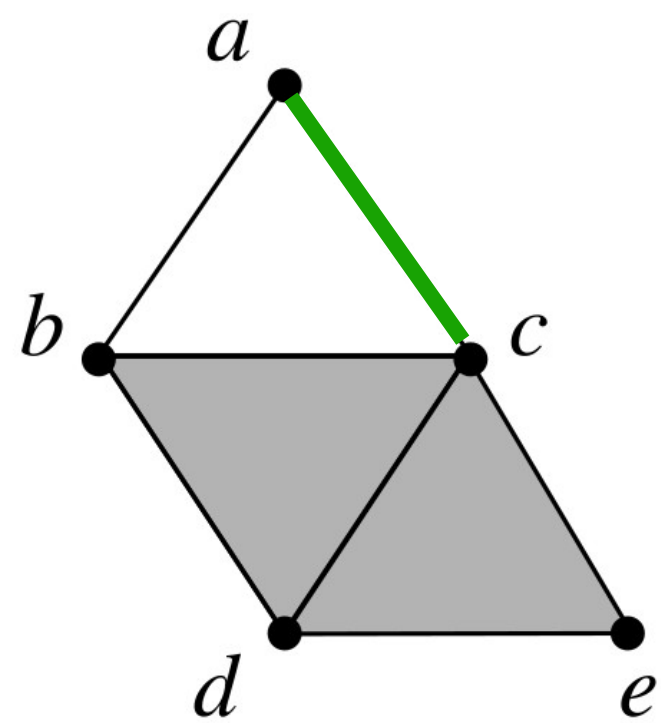
  satisfying a **compatibility condition**

$$
\begin{array}{c}
 \\ a \\ b \\ c \\ d \\ e
\end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 1
\end{pmatrix}
\end{array}
\xrightarrow{g,\ f}
\begin{array}{c}
\begin{array}{ccccc} g(1) & g(2) & g(3) & g(4) & g(5) \end{array} \\
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1
\end{pmatrix}
\end{array}
\begin{array}{c} \\ A \\ B \\ C \end{array}
$$

Classic rep'n: Dowker complex

- Each row specifies a vertex
- Each column specifies (at least one) simplex by selecting subsets of vertices

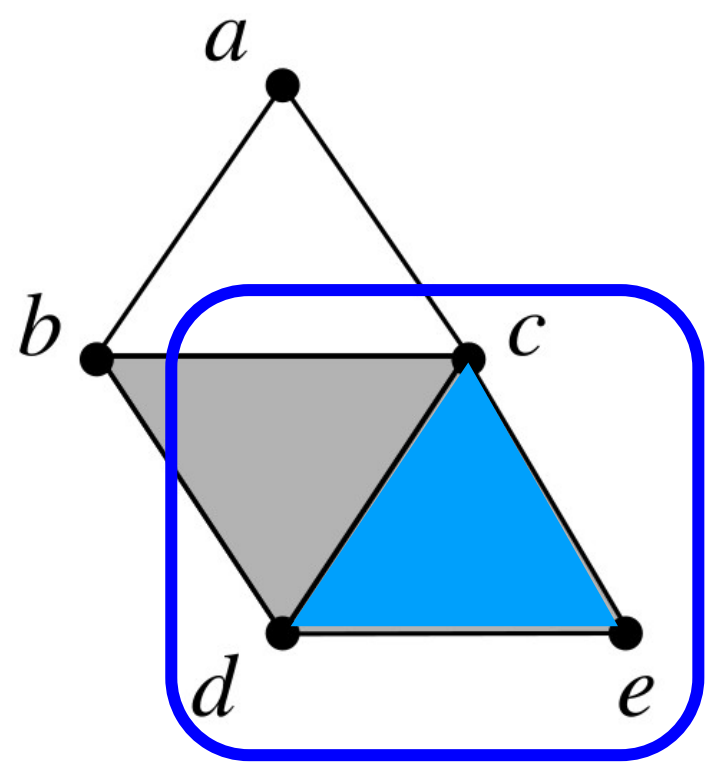
$$
\begin{array}{c c}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} & \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}
\end{array}
$$



Classic rep'n: Dowker complex

- Each row specifies a vertex
- Each column specifies (at least one) simplex by selecting subsets of vertices

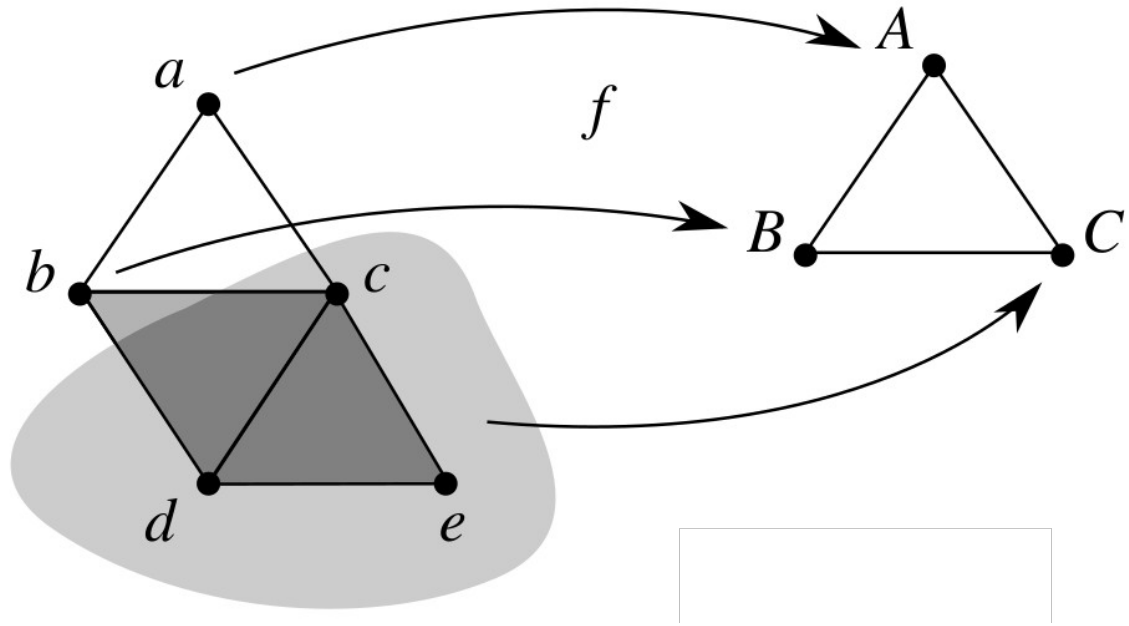
$$
\begin{array}{c c}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} & \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}
\end{array}
$$



The Dowker complex is functorial

<u>Theorem</u>: Functoriality for inclusions of relations (Chowdhury and Mémoli, *JACT*, 2018.)

<u>Theorem</u>: Also true at full generality!

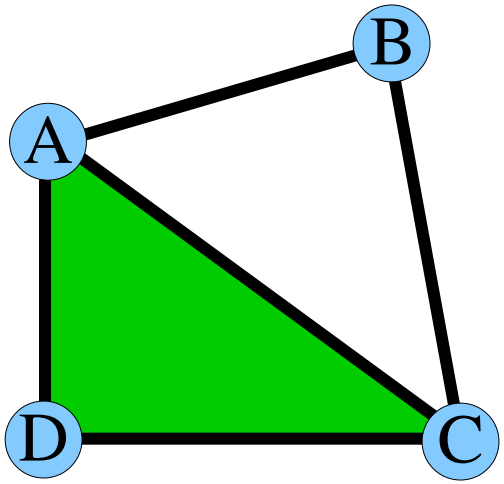


$$\begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \xrightarrow{f} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{array}{c} A \\ B \\ C \end{array}$$

Dowker is lossy

- Dowker ignores duplicate columns
- Here are several non-isomorphic relations inducing the same complex

$$\text{parsers}\quad \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \text{files} \\ \begin{pmatrix} 100000111111000011111 \\ 011000111000100000000 \\ 000110000100111111111 \\ 000001000011011111111 \end{pmatrix} \end{array}$$



$$\text{parsers}\quad \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \text{files} \\ \begin{pmatrix} 101001 \\ 110000 \\ 011101 \\ 001010 \end{pmatrix} \end{array}$$

$$\text{parsers}\quad \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \text{files} \\ \begin{pmatrix} 00110110 \\ 01100000 \\ 01011000 \\ 00010101 \end{pmatrix} \end{array}$$

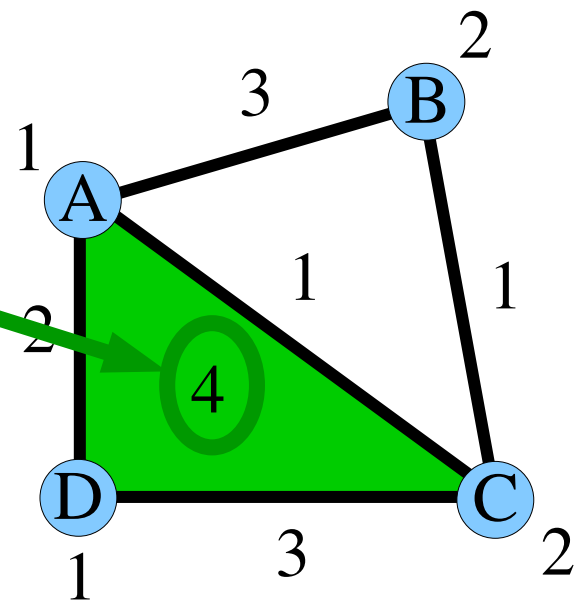
Weighted Dowker complex

- Weighting: Count how many times each simplex appears
- <u>Theorem</u>: The matrix is determined (up to isomorphism) by the Dowker complex with this weight function
- <u>Deeper theorem</u>: This can be enriched into a covariant functor

files

parsers

$$\begin{array}{c} A \\ B \\ C \\ D \end{array} \begin{pmatrix} 1000001111110000\boxed{1111} \\ 0110001110001000\boxed{0000} \\ 0001100001001111\boxed{1111} \\ 0000010000110111\boxed{1111} \end{pmatrix}$$

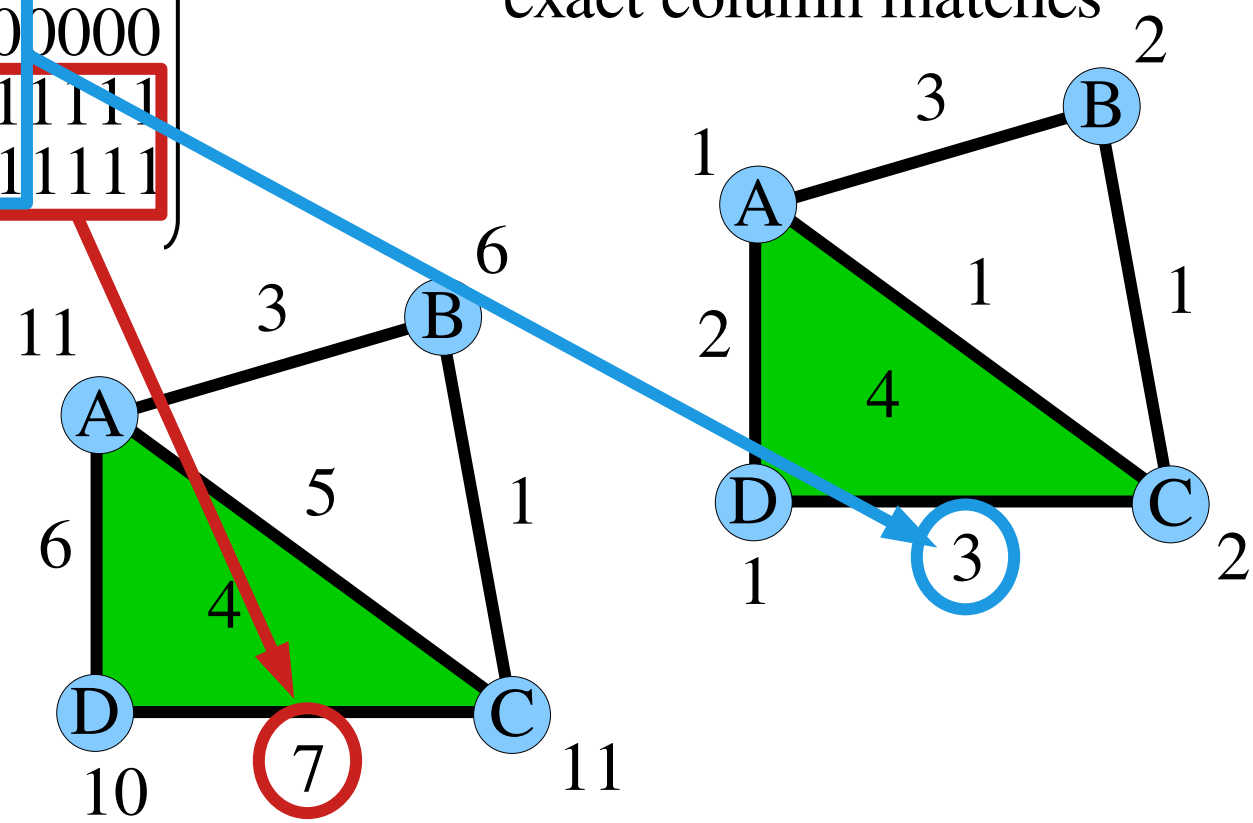


Dowker weight functions

- There are actually **two** weight functions that distinguish isomorphism classes

$$\text{parsers}\quad \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \overset{\text{files}}{\begin{pmatrix} 1000001111110001111 \\ 0110001110001000000 \\ 0001100001001111111 \\ 0000010000110111111 \end{pmatrix}}$$

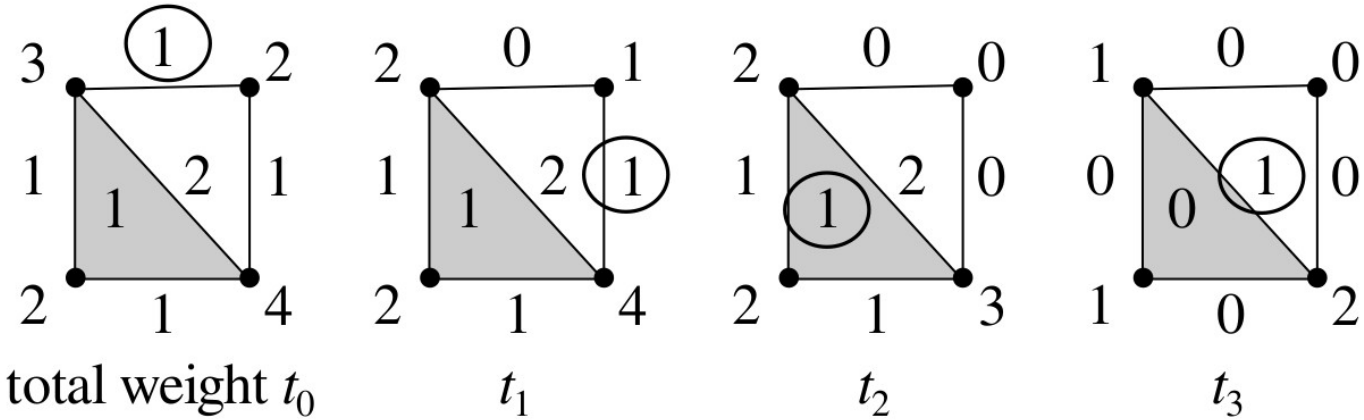
Differential weight:
For each simplex, count exact column matches



Total weight:
For each simplex, count columns matching the vertices present. Ignore the other vertices

Dowker weight functions

- <u>Theorem</u>: Both total and differential weight are complete isomorphism invariants for **Rel**.
- The proof is constructive and algorithmic



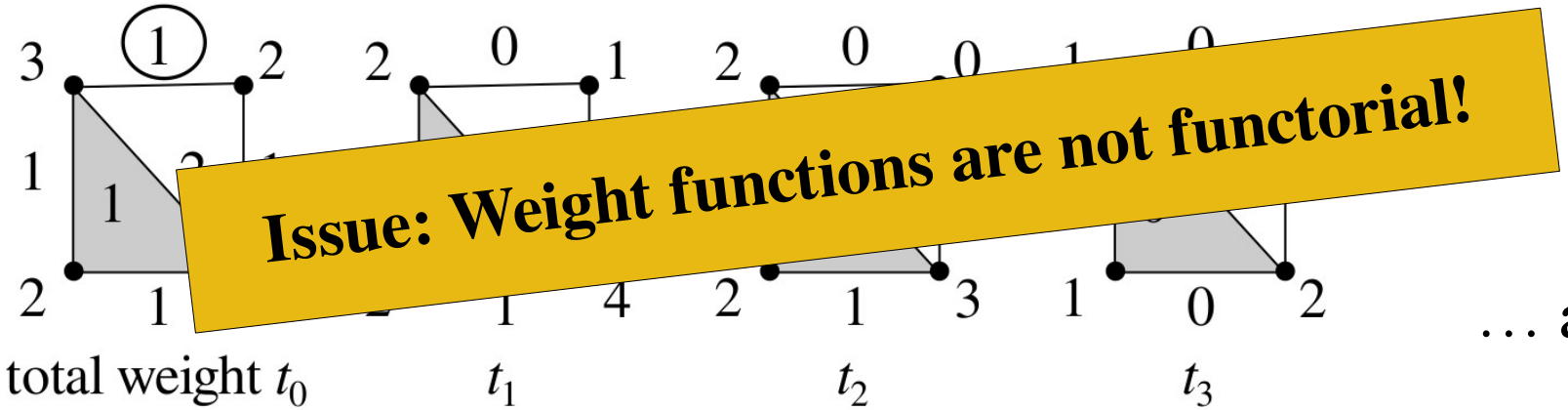
total weight t_0 t_1 t_2 t_3

… and so on!

$$r_0 = \begin{pmatrix} \end{pmatrix} \qquad r_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad r_2 = \begin{pmatrix} 10 \\ 11 \\ 01 \\ 00 \end{pmatrix} \qquad r_3 = \begin{pmatrix} 101 \\ 110 \\ 011 \\ 001 \end{pmatrix}$$

Dowker weight functions

- <u>Theorem</u>: Both total and differential weight are complete isomorphism invariants for **Rel**.
- The proof is constructive and algorithmic



Issue: Weight functions are not functorial!

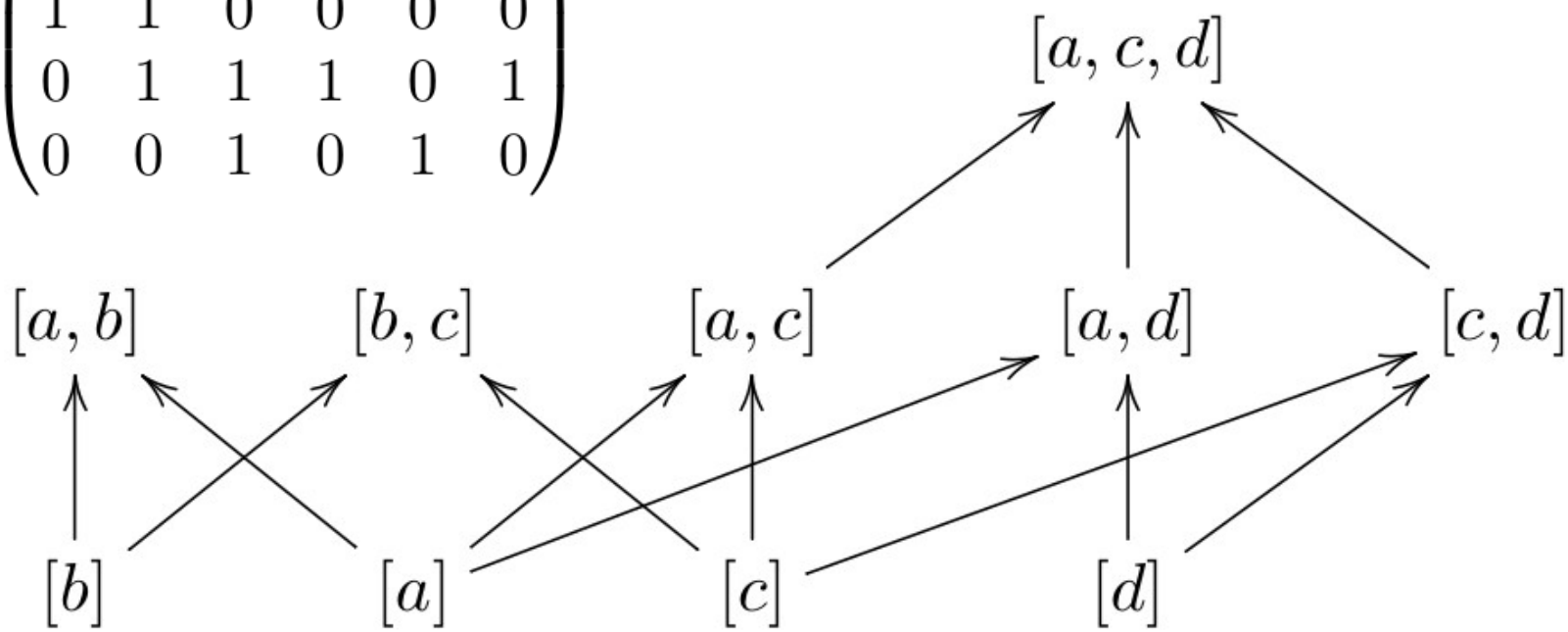
… and so on!

$$r_0 = \begin{pmatrix} \end{pmatrix} \qquad r_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad r_2 = \begin{pmatrix} 10 \\ 11 \\ 01 \\ 00 \end{pmatrix} \qquad r_3 = \begin{pmatrix} 101 \\ 110 \\ 011 \\ 001 \end{pmatrix}$$

Reorganizing the data using posets

- Draw the Hasse diagram for face poset of the Dowker complex (arrows indicate simplicial inclusions)

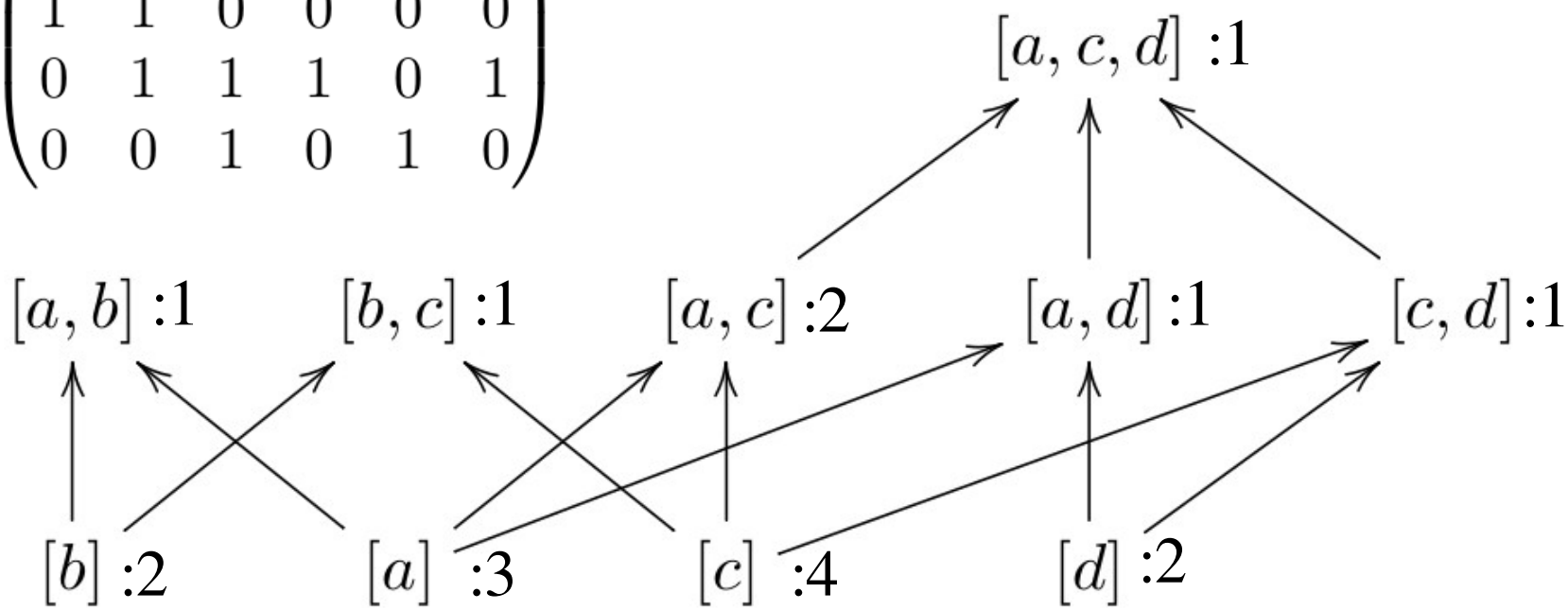
$$
\begin{array}{cc}
 & \{1,\ 2,\ 3,\ 4,\ 5,\ 6\} = Y \\
X = \begin{array}{c} \{a, \\ b, \\ c, \\ d\} \end{array} & \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}
\end{array}
$$



Reorganizing the data using posets

- Label the directed graph with total* weights for a lossless representation up to **Rel** isomorphism

$$
\begin{array}{cc}
 & \{1,\ 2,\ 3,\ 4,\ 5,\ 6\} = Y \\
X = \begin{array}{c}\{a,\\ b,\\ c,\\ d\}\end{array} & \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}
\end{array}
$$



*Differential weights seem to be more useful in practice, though…

Dowker cosheaf representation

- But instead of total weights, we can just list the columns!

$$X = \{a, b, c, d\} \quad \begin{array}{c} \{1,\ 2,\ 3,\ 4,\ 5,\ 6\} = Y \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

If we want the arrows to be functions, we have to reverse them… They become set inclusions.
The diagram is actually a *cosheaf*

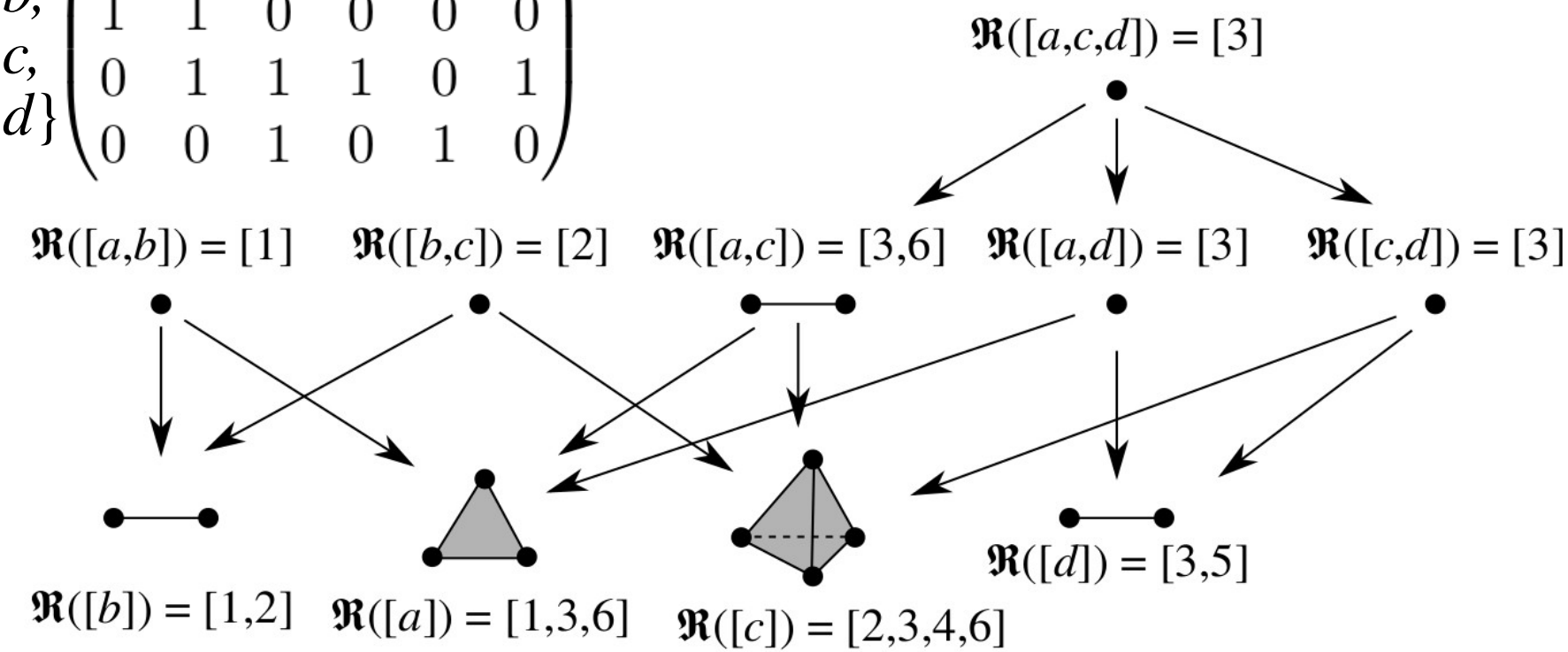
$$\begin{tikzcd}
 & & \mathfrak{R}^0([a,c,d])= \{3\} \arrow[dl] \arrow[d] \arrow[dr] & & \\
\mathfrak{R}^0([a,b])= \{1\} \arrow[d] \arrow[dr] & \mathfrak{R}^0([b,c])= \{2\} \arrow[dl] \arrow[dr] & \mathfrak{R}^0([a,c])= \{3,6\} \arrow[d] \arrow[dl] & \mathfrak{R}^0([a,d])= \{3\} \arrow[d] \arrow[dll] & \mathfrak{R}^0([c,d])= \{3\} \arrow[dl] \arrow[dll] \\
\mathfrak{R}^0([b])= \{1,2\} & \mathfrak{R}^0([a])= \{1,3,6\} & \mathfrak{R}^0([c])= \{2,3,4,6\} & \mathfrak{R}^0([d])= \{3,5\} &
\end{tikzcd}$$

Dowker cosheaf representation

- Each *costalk* is actually a simplex, giving a cosheaf of simplicial complexes on a simplicial complex!

Notation:
CoShvAsc = category of these cosheaves

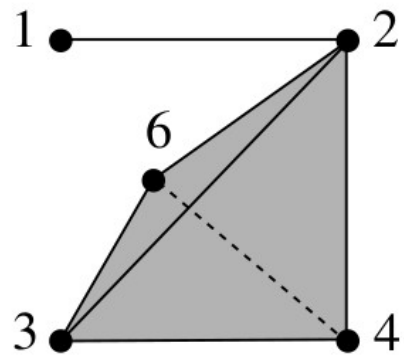
$$X = \{a, b, c, d\} \quad \begin{array}{c} \{1,\ 2,\ 3,\ 4,\ 5,\ 6\} = Y \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

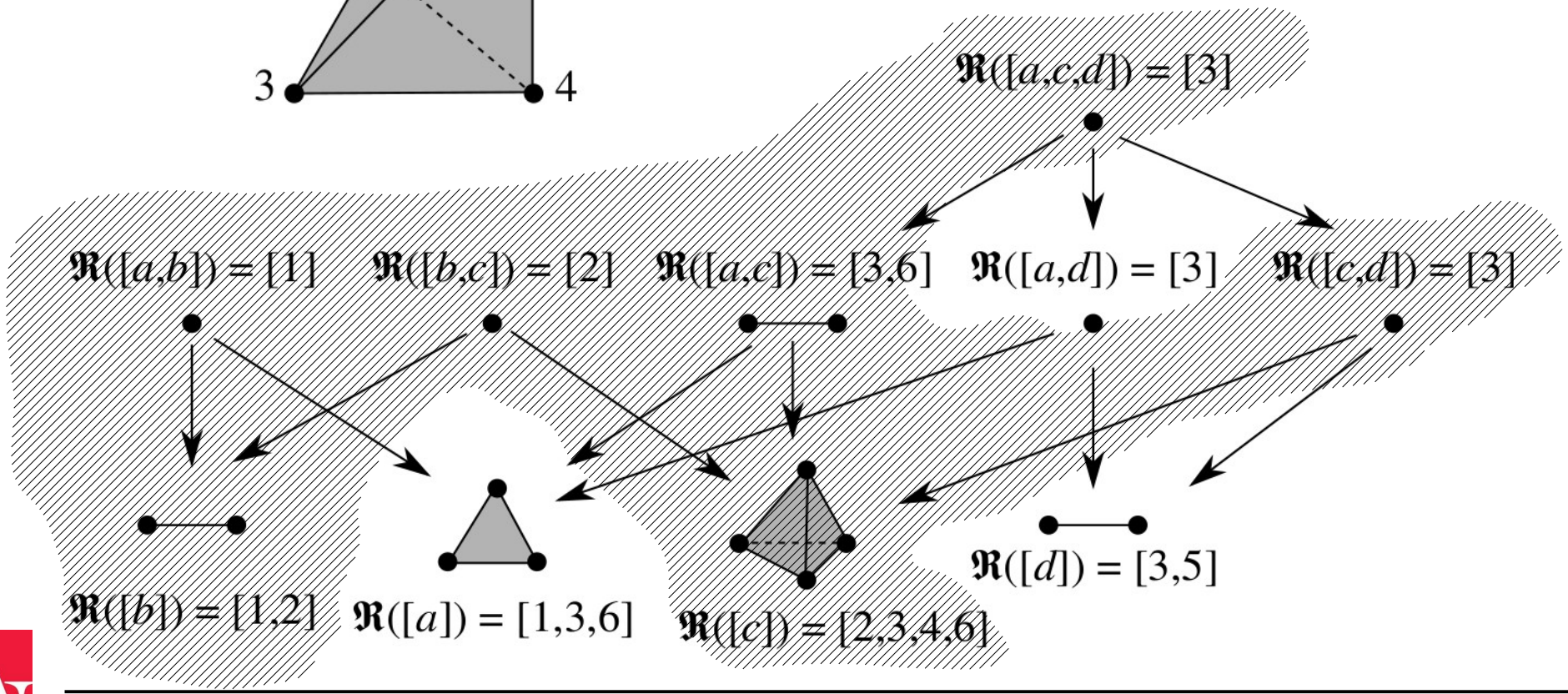


Cosections may be arbitrary complexes

- Cosections on any *open set* are computed by gluing

(upwardly closed in the Hasse diagram = downwardly closed in the cosheaf diagram)

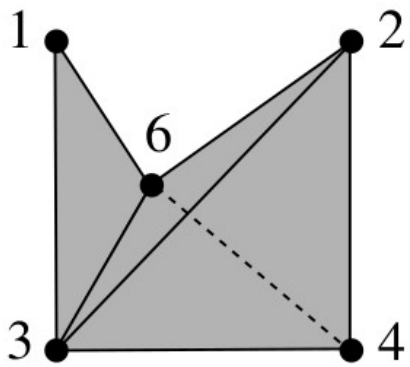


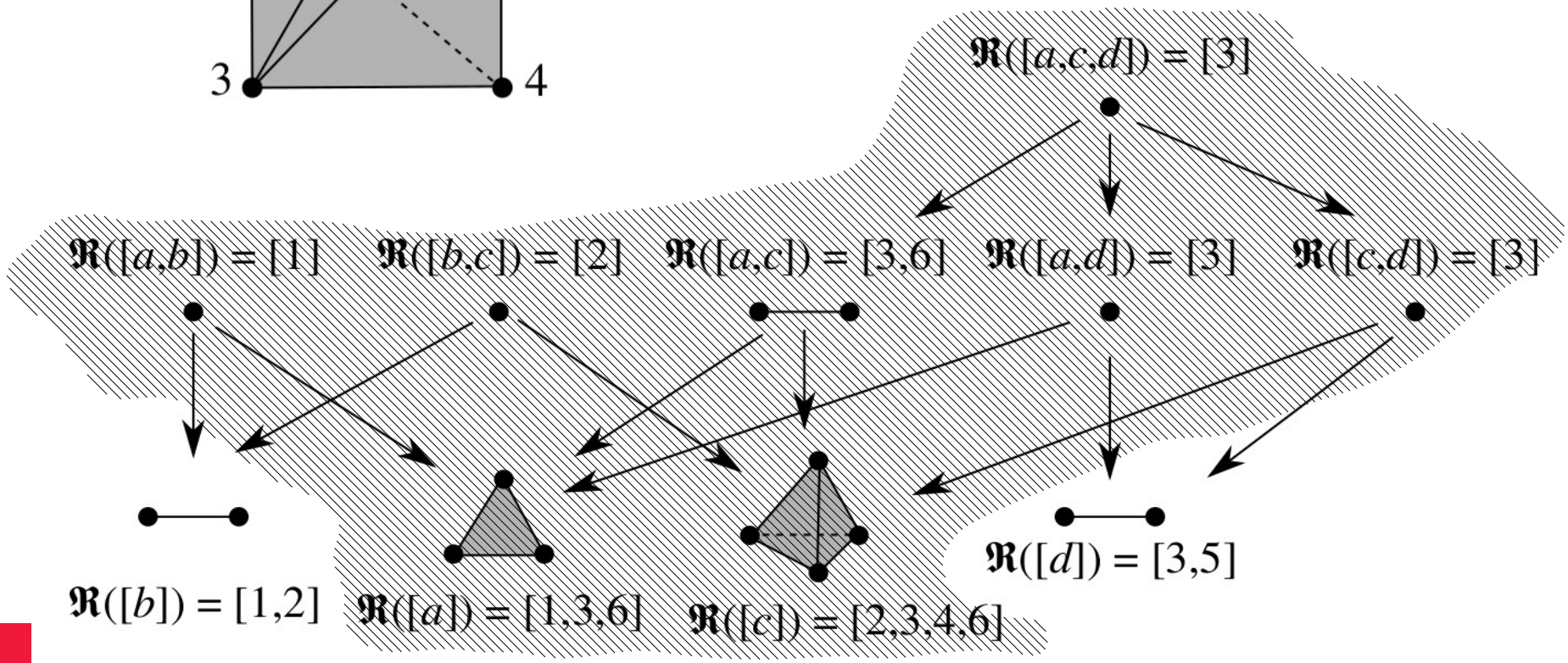


Cosections may be arbitrary complexes

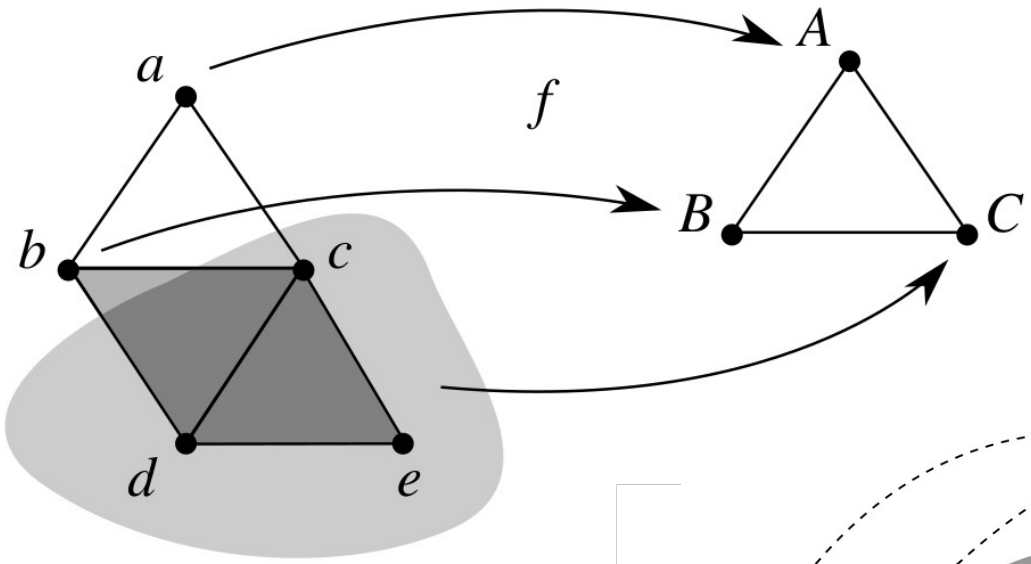
- Cosections on any *open set* are computed by gluing

(upwardly closed in the Hasse diagram = downwardly closed in the cosheaf diagram)



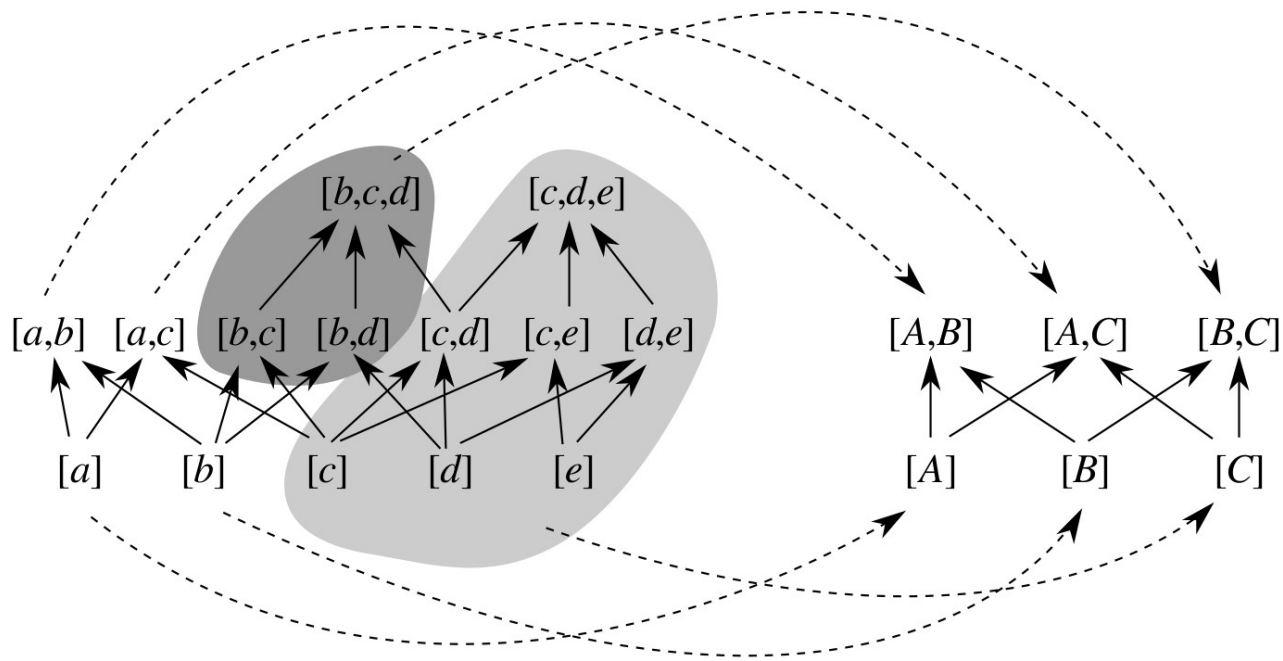


The Dowker complex is functorial...



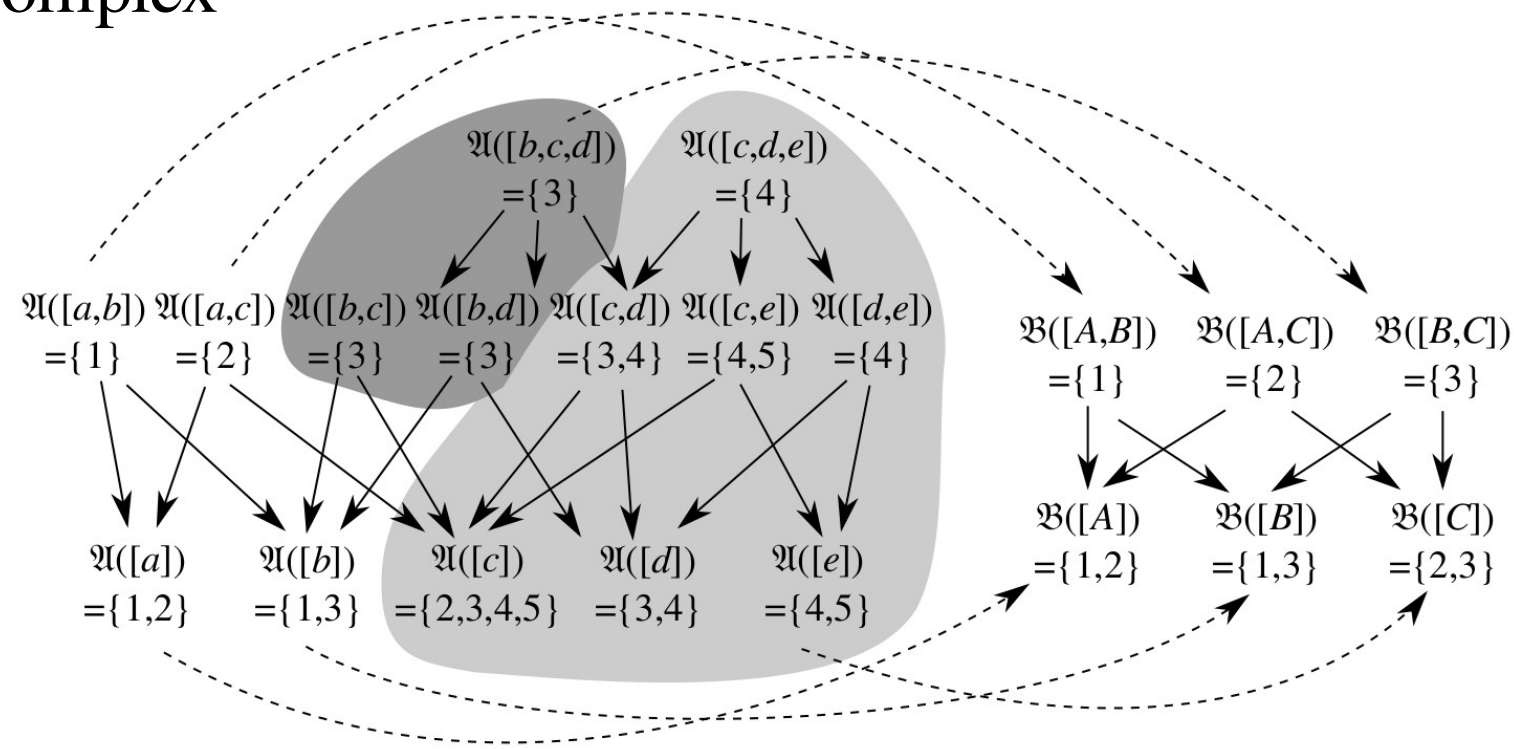
Face partial orders are more useful for most things than simplicial complexes!

For some mysterious reason, this appears to be a folk theorem (?!) It's not hard; exercise for the audience...



… and so is the Dowker cosheaf

- <u>Theorem</u>: The Dowker cosheaf construction defines a faithful covariant functor* **Rel** → **CoShvAsc**.
- The proof mostly follows the functoriality proof for the Dowker complex



*Most authors don't let the base space vary in a (co)sheaf morphism. Beats me why! It's usually better to allow base space maps, and I need them here.

Dowker duality

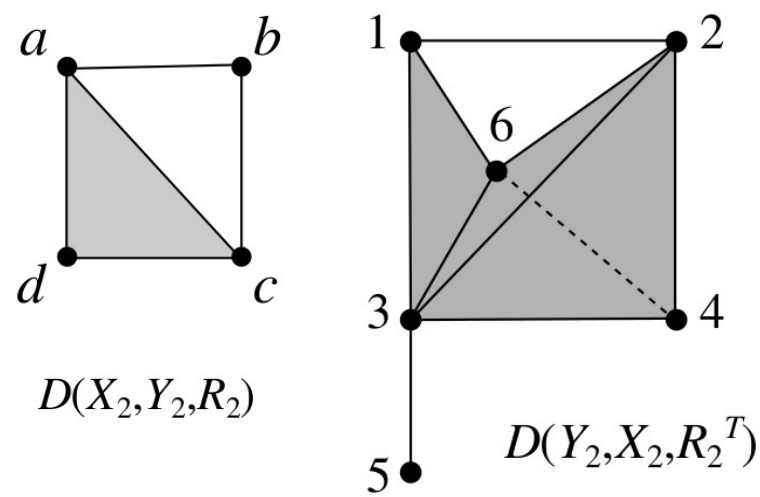
- Dowker's name is attached to these constructions because…
- <u>Theorem:</u> (Dowker*, 1952)

$$H_\bullet(D(X,Y,R)) \cong H_\bullet(D(Y,X,R^T))$$

(The matrix of R^T is the transpose of the matrix of R)

- This has been strengthened to a homotopy equivalence
- Several other generalizations are known

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$



$D(X_2,Y_2,R_2)$

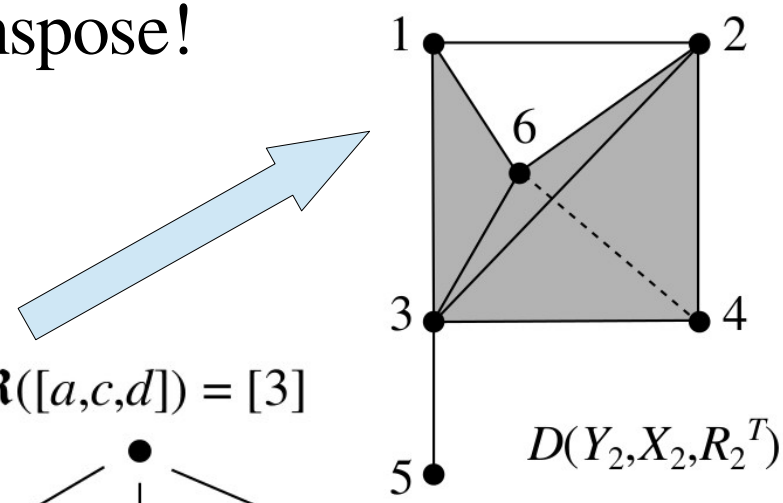
$D(Y_2,X_2,R_2^T)$

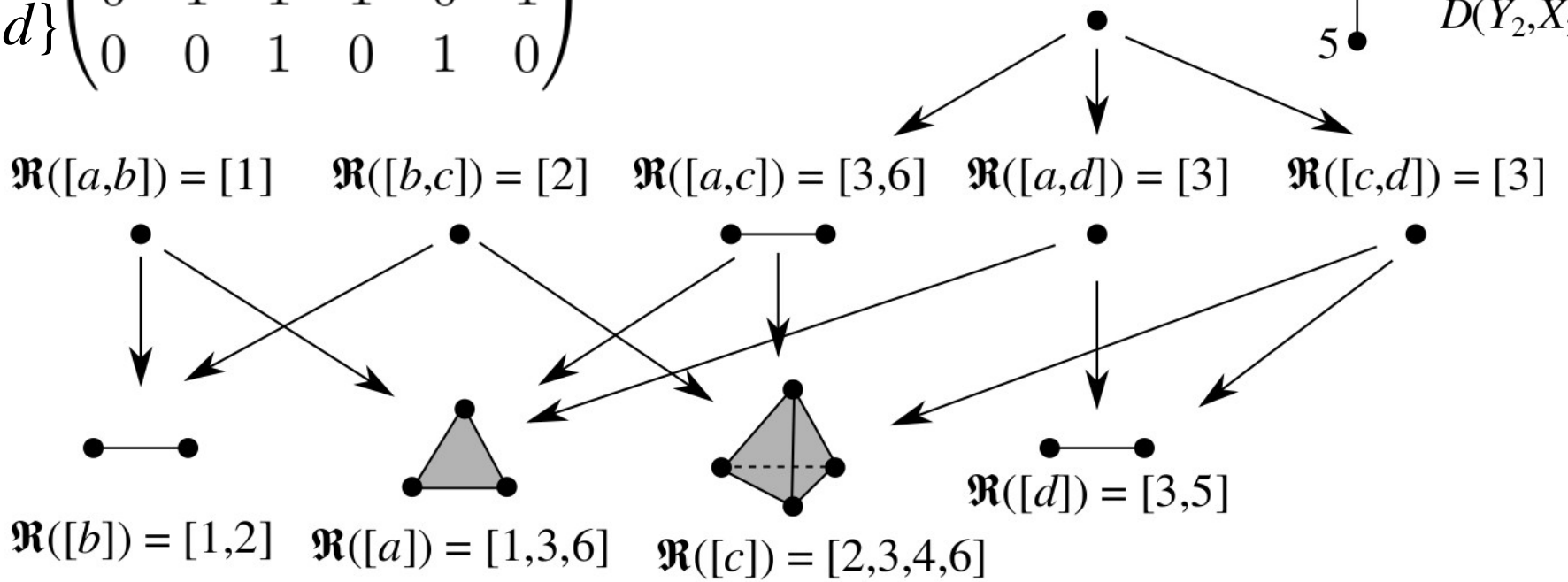
*I found Dowker's paper **very** rough going.
Chowdhury & Mémoli's paper is much easier!

An interesting observation

- The space of global cosections for the Dowker cosheaf is itself a Dowker complex… of the transpose!

$$X = \begin{array}{c} \\ \{a, \\ b, \\ c, \\ d\} \end{array} \begin{array}{c} \{1,\ 2,\ 3,\ 4,\ 5,\ 6\} = Y \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{array}$$



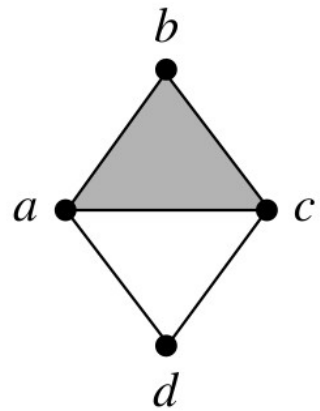


An interesting observation

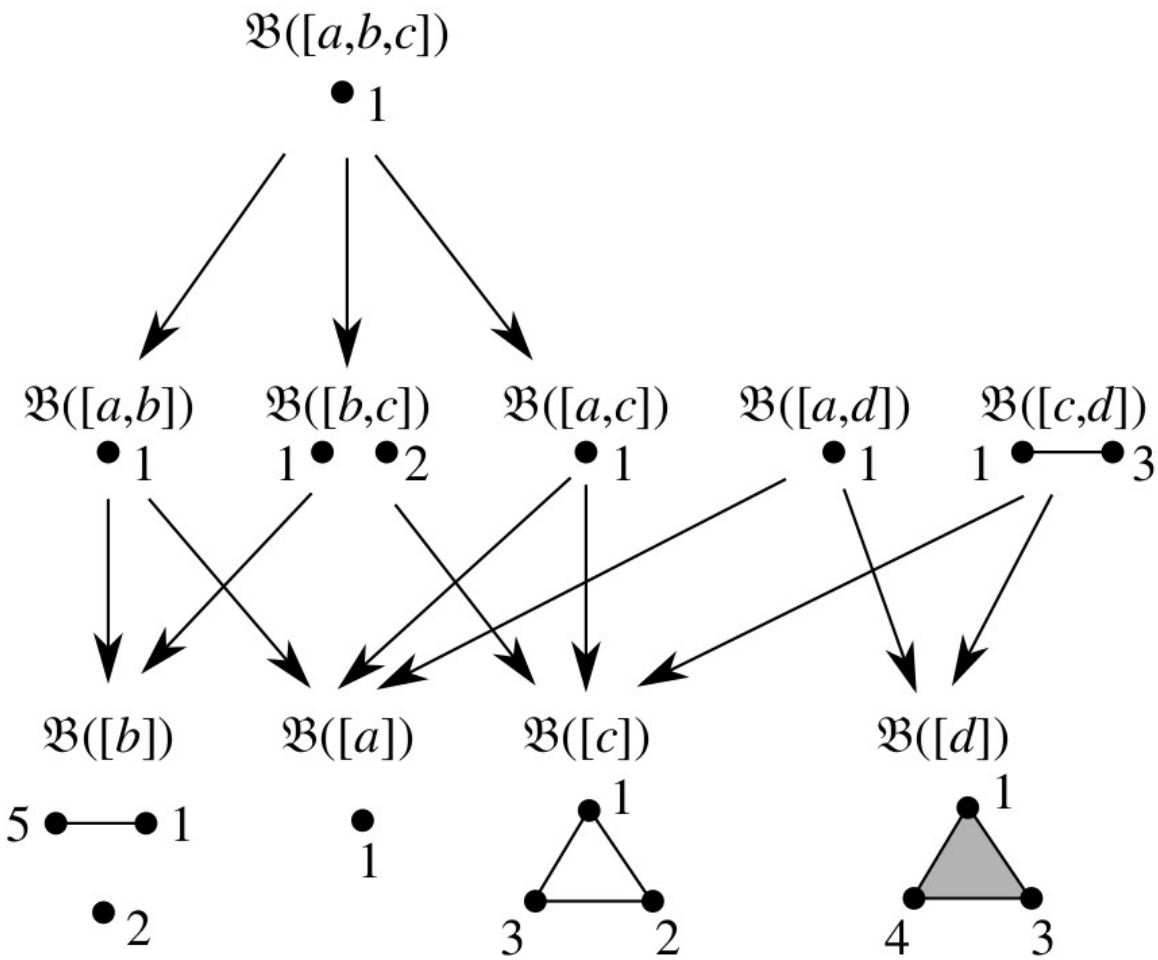
- The space of global cosections for the Dowker cosheaf is itself a Dowker complex… of the transpose!
- The Dowker cosheaf therefore has both the Dowker complex (base space) and its dual (global cosections) baked into it.
- Not only that, this property is **functorial**!
- <u>Theorem</u>: There is a duality functor *Dual* : **CoShvAsc** → **CoShvAsc** that exchanges the base space with the cosections
- The proof, like the others, is an elaborate diagram construction
- Moreover, the definition of *Dual* works for all cosheaves of abstract simplicial complexes

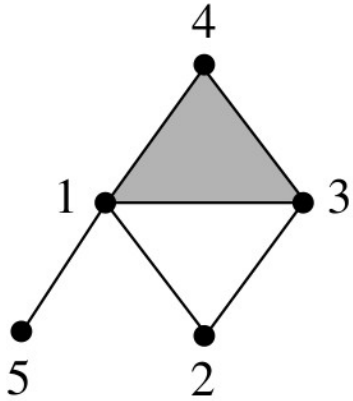
Cosheaf-of-ASC duality functor

- The *Dual* functor acts on non-Dowker cosheaves as well



Base space





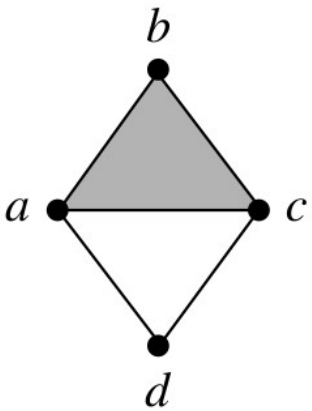
Global cosections

Some random cosheaf of abstract simplicial complexes

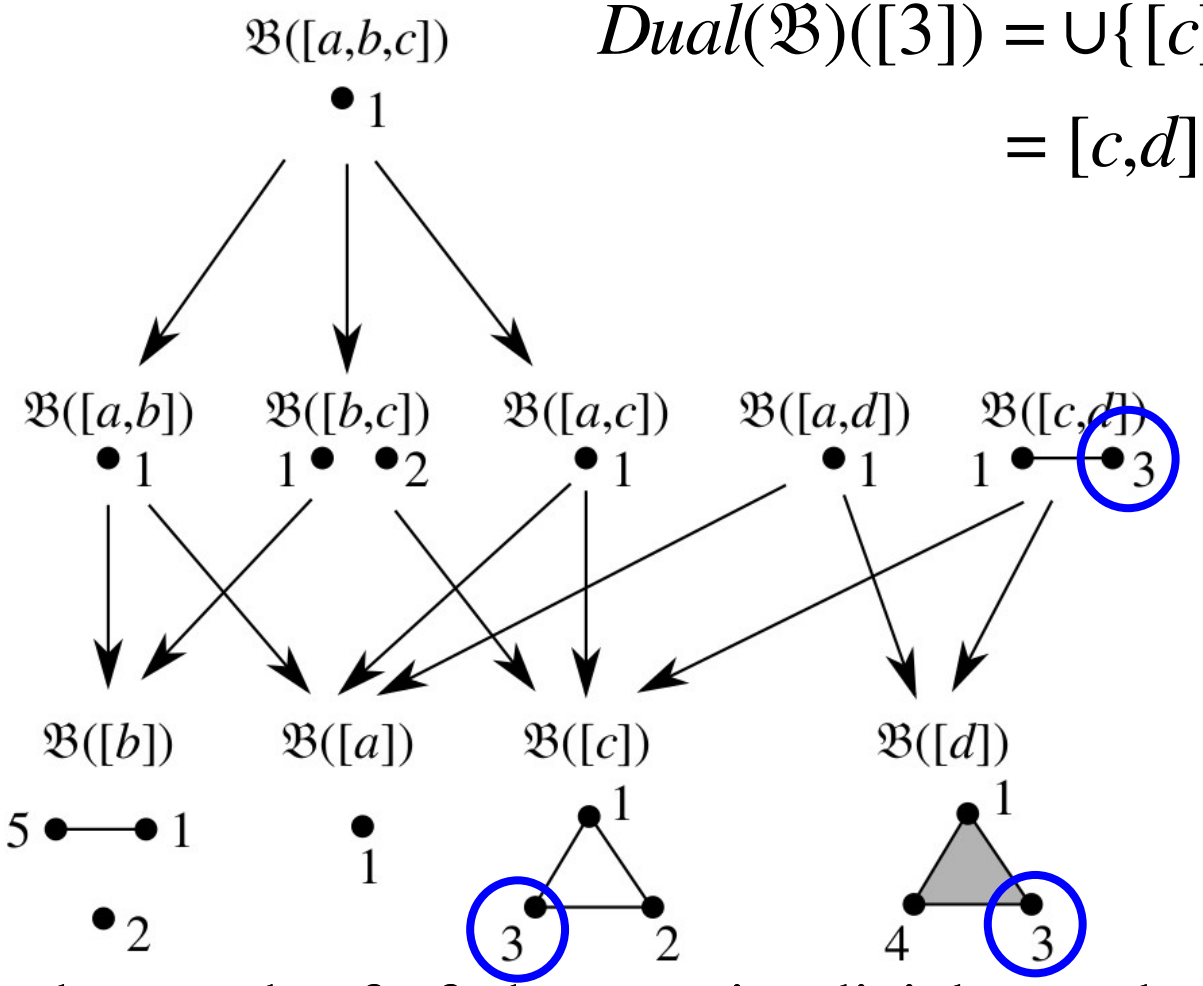
Cosheaf-of-ASC duality functor

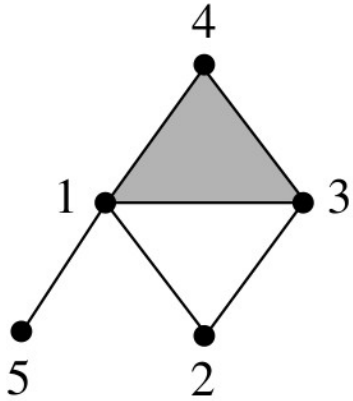
- $Dual(\mathfrak{B})(\sigma)$:= union of all simplices α whose costalks $\mathfrak{B}(\alpha)$ contain σ

$$Dual(\mathfrak{B})([3]) = \cup\{[c],[d],[c,d]\}$$
$$= [c,d]$$



Base space





Global cosections

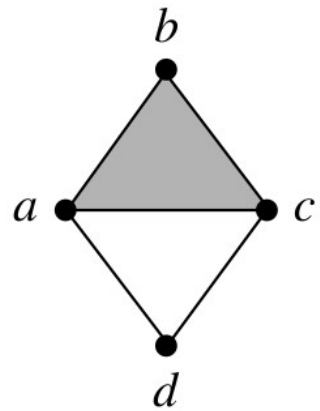
Some random cosheaf of abstract simplicial complexes

Cosheaf-of-ASC duality functor

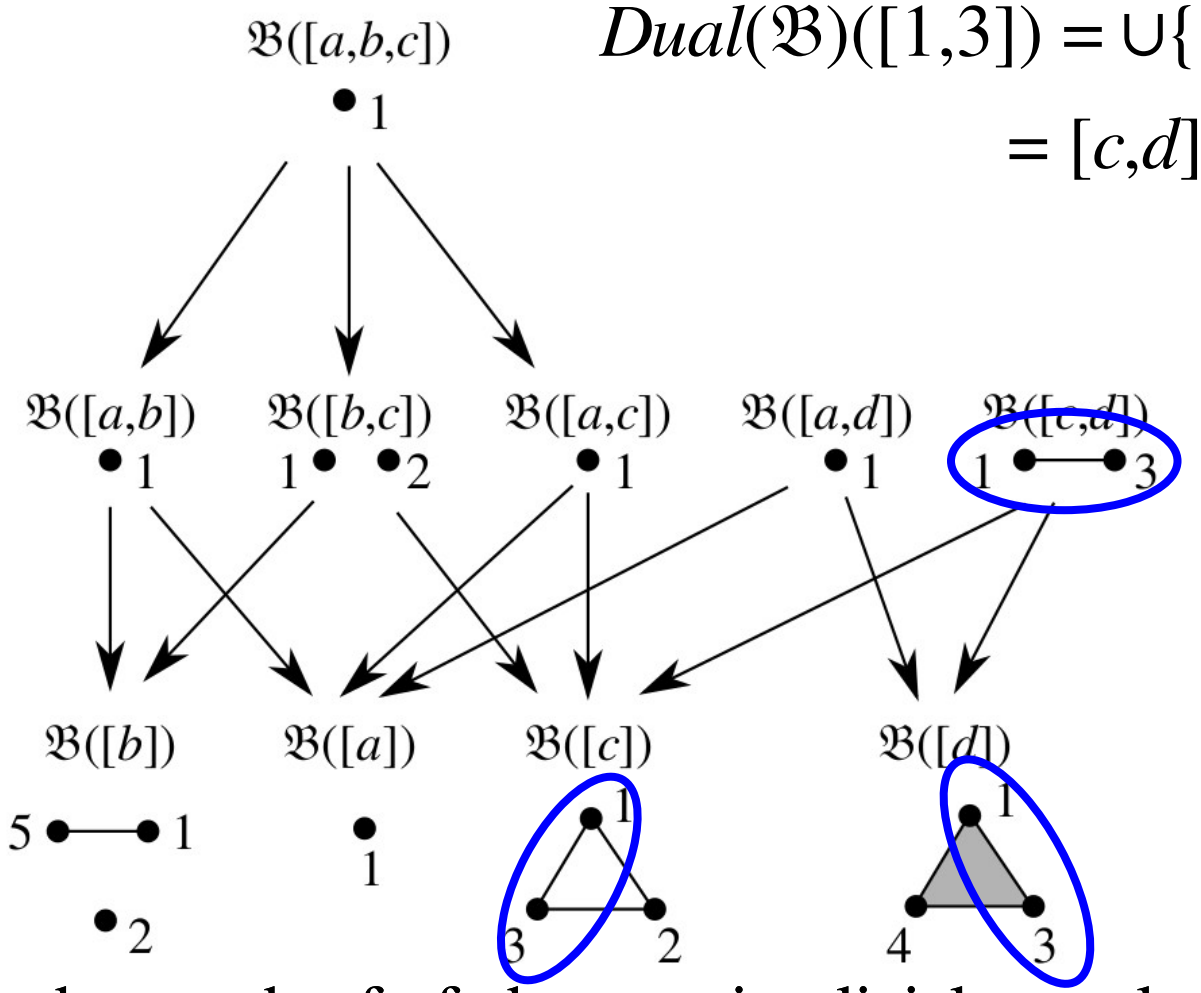
- $Dual(\mathfrak{B})(\sigma)$:= union of all simplices α whose costalks $\mathfrak{B}(\alpha)$ contain σ

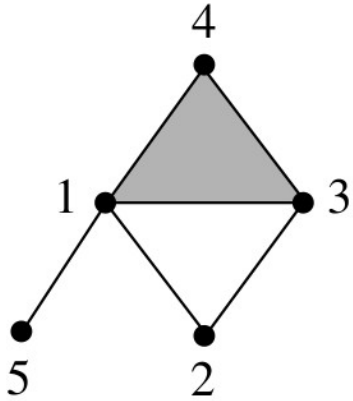
$Dual(\mathfrak{B})([1,3]) = \cup\{[c],[d],[c,d]\}$

$= [c,d]$



Base space



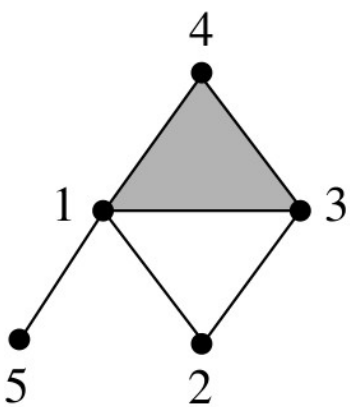


Global cosections

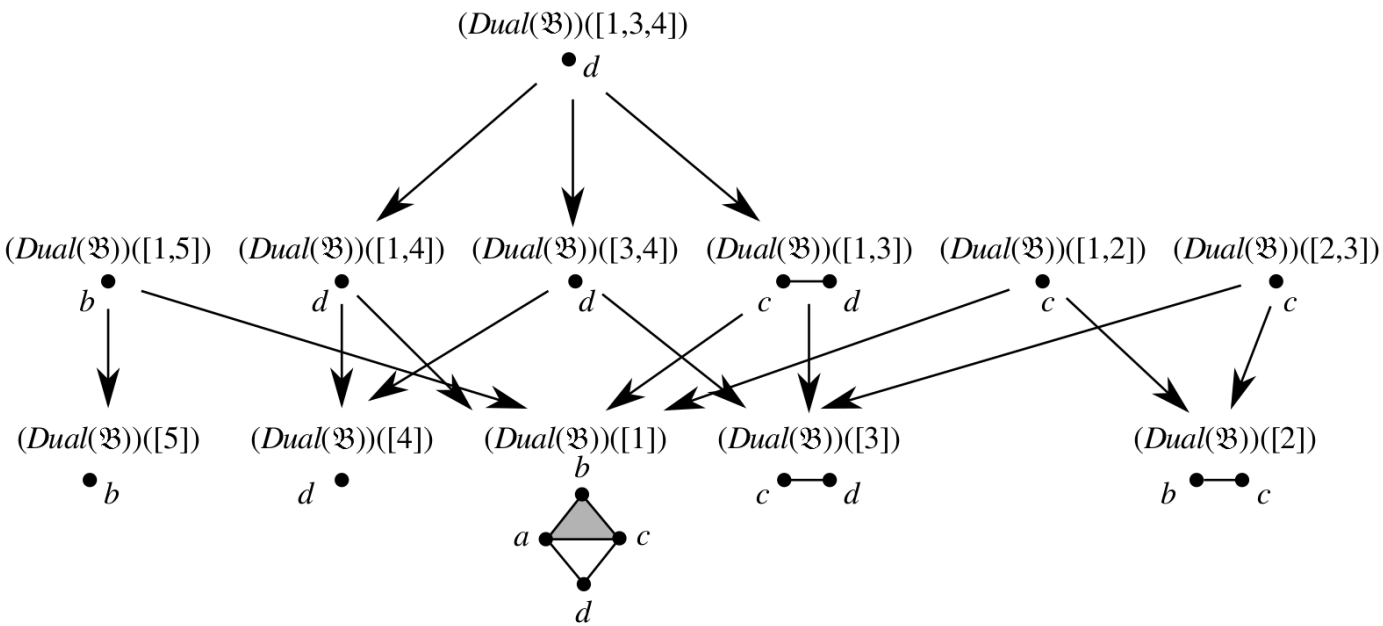
Some random cosheaf of abstract simplicial complexes

Cosheaf-of-ASC duality functor

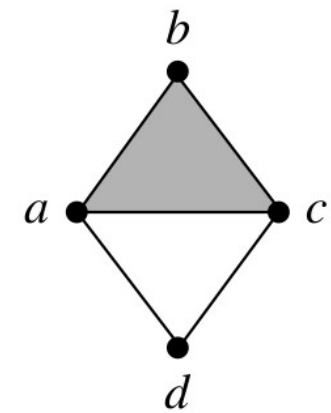
- *Dual* exchanges the base space and space of global cosections
- Note that costalks might not be complete simplices if we aren't working with Dowker cosheaves



Base space



Dual cosheaf



Global cosections

Cosheaf Dowker duality

- <u>Theorem</u>: The space of cosections of the Dowker cosheaf is the Dowker dual of its base space.
  - Briefly, the following functor diagram commutes:

$$\begin{tikzcd}
\mathbf{Rel} \arrow[r, "CoShvRep"] \arrow[d, "Transp"'] & \mathbf{CoShvAsc} \arrow[d, "Dual"] \\
\mathbf{Rel} \arrow[r, "CoShvRep"'] & \mathbf{CoShvAsc}
\end{tikzcd}$$

  - The only thing that needs to be shown is that dualizing a Dowker cosheaf yields a new cosheaf whose costalks are all complete simplices … and a bit of calculation besides.

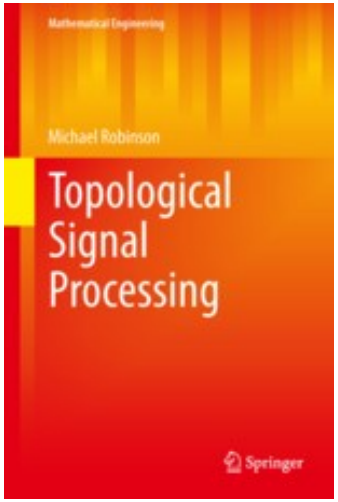
Implications for consensus file formats

- Because of functoriality, any of these Dowker constructions are reasonable tools for studying parser behavior on files
- Functoriality guides the process of summarization
  - Which files are good exemplars of (non)compliance?
  - Which parsers are redundant, or conversely, which have divergent capabilities?
  - Statistics on the values of the weights “makes sense”
- Functoriality guides the process of curating corpora
  - If we test the same set of parsers on two different corpora, how do we compare the results?
- You can consider parser-file relations or file-parser relations without losing anything

Next steps

- We can generalize **Rel** to take poset values rather than booleans
  - This breaks most of the existing technology for filtrations, but **not** our cosheaf constructions
- <u>Theorem</u>: The Dowker cosheaf generalized for poset-valued **Rel** is functorial
  - What can be said about persistence constructions in this case?
- We recently discovered a pseudometric on **Rel**, which gives it a topology
- <u>Theorem</u>: The total and differential weight functions are both continuous with respect to this topology on **Rel** under the sup-norm
  - Chowdhury & Mémoli's starting point is a filtration on the Dowker complex, which the total weight provides
  - This suggests that *interleavings* of cosheaves might work. (Cool! I have some results on sheaf-based general interleavings)
- Explore more of the category theoretic structure of **Rel**
  - I suspect it has at least two distinct (co)products and more category theoretic delights!

To learn more...



Michael Robinson

michaelr@american.edu

http://drmichaelrobinson.net

Relevant preprints:

https://arxiv.org/abs/2003.00976

https://arxiv.org/abs/2005.12348

Software:

https://github.com/kb1dds