ALGODAN Algorithmic Data Analysis

The Sum-Product Bridge

Mikko Koivisto
Academy Research Fellow

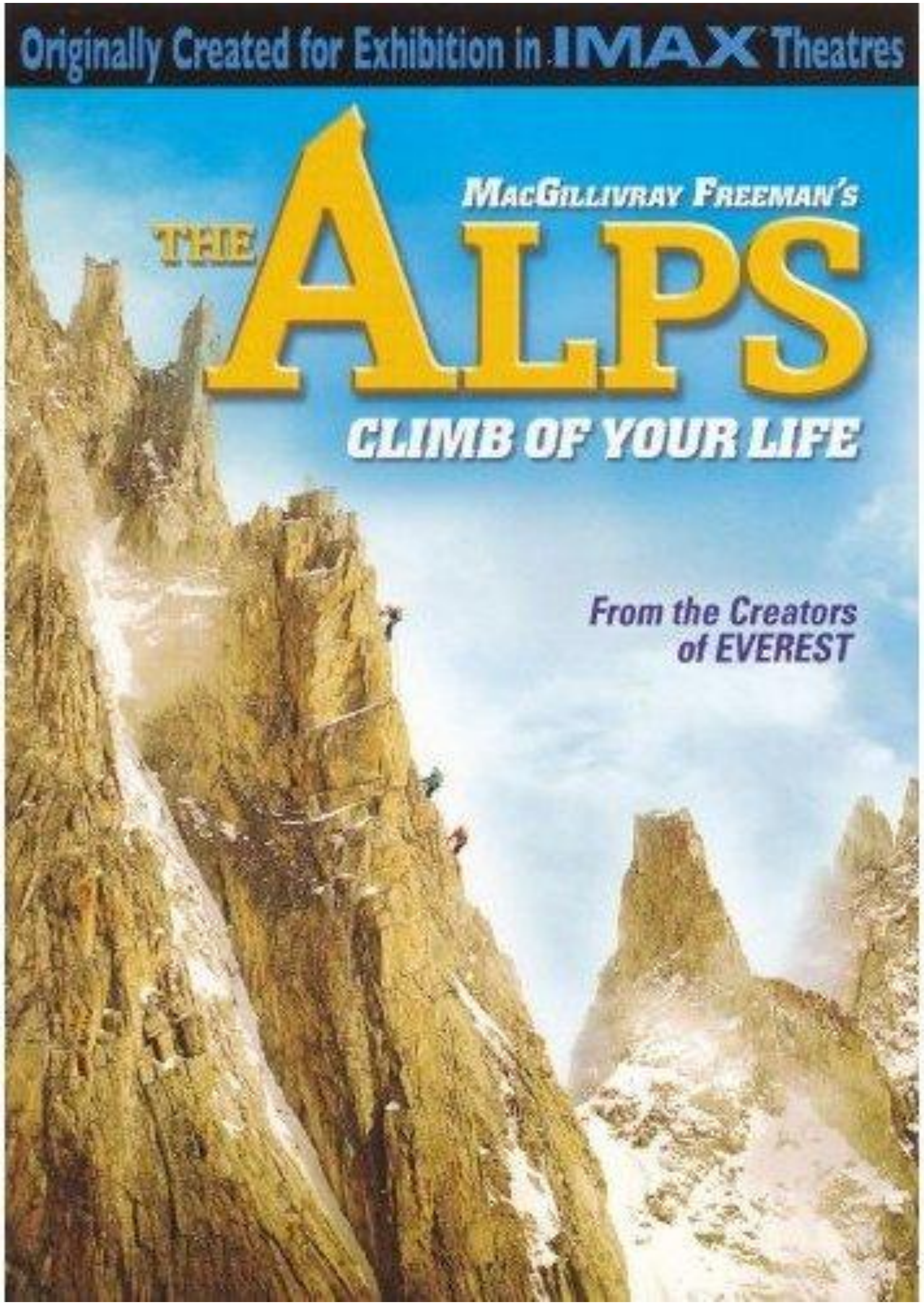
Aalto University

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Sums of Products United – People

- Club members at the CS department of UH
  - Esther Galbrun, *PhD student (co-advisor H. Toivonen)*
  - Kustaa Kangas, *MSc student*
  - Mikko Koivisto, *leader, advisor*
  - Janne Korhonen, *PhD student (co-advisor P. Kaski)*
  - Teppo Niinimäki, *PhD student*
  - Pekka Parviainen, *PhD student*
- Associate club members (key collaborators)
  - Petteri Kaski (Algodan, Aalto University)
  - Andreas Björklund (Lund University)
  - Thore Husfeldt (IT University Copenhagen)
  - ...

ALGODAN Algorithmic Data Analysis


Originally Created for Exhibition in IMAX Theatres
MacGillivray Freeman's
THE ALPS
CLIMB OF YOUR LIFE
From the Creators of EVEREST

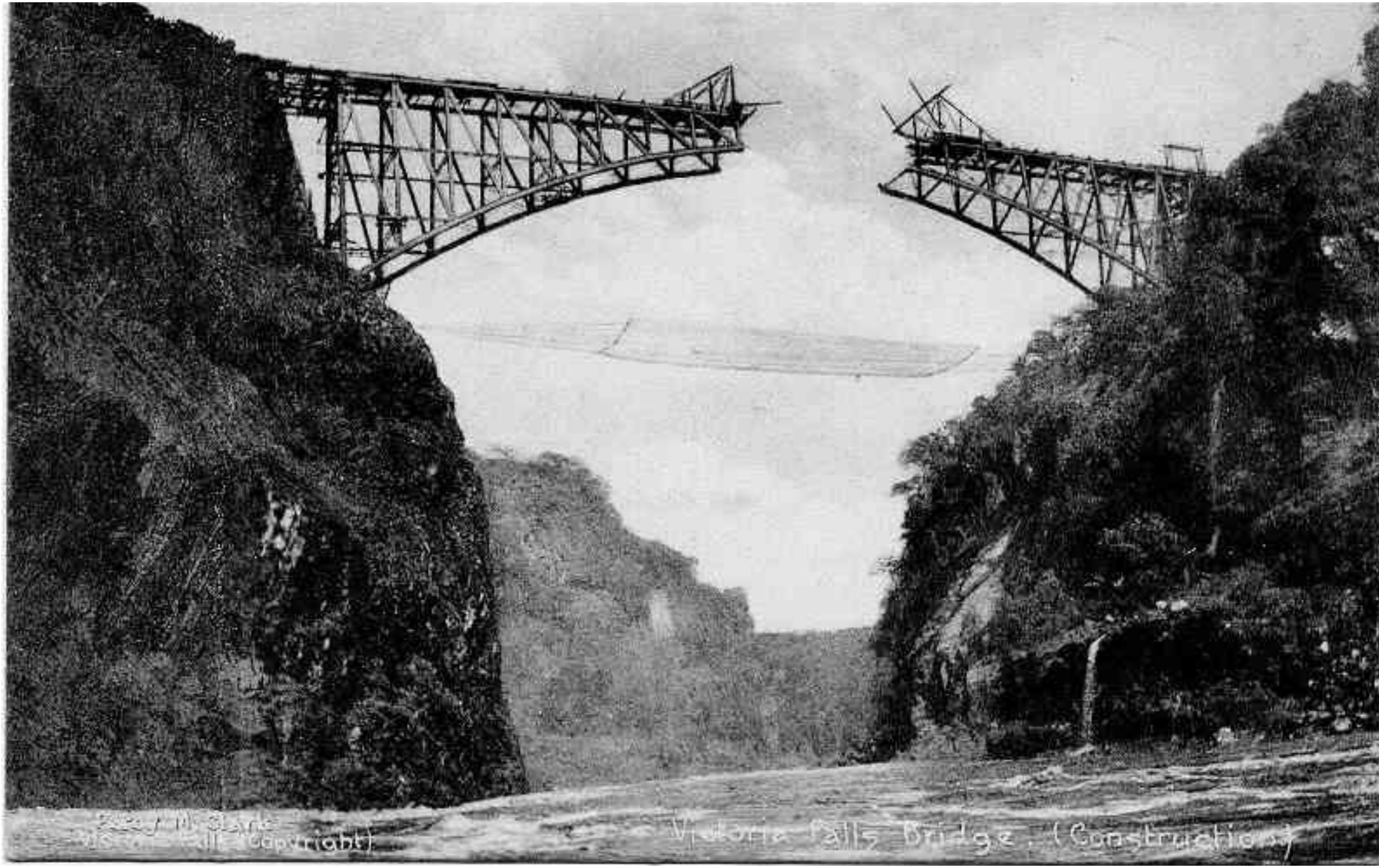
SoPU – Mission

Build and maintain a bridge that connects algorithm theory and computational statistics by developing the methodology of computing large sums of products.

The amazing Fairyland Bridge connects two mountains at 5,000ft at Huangshan.

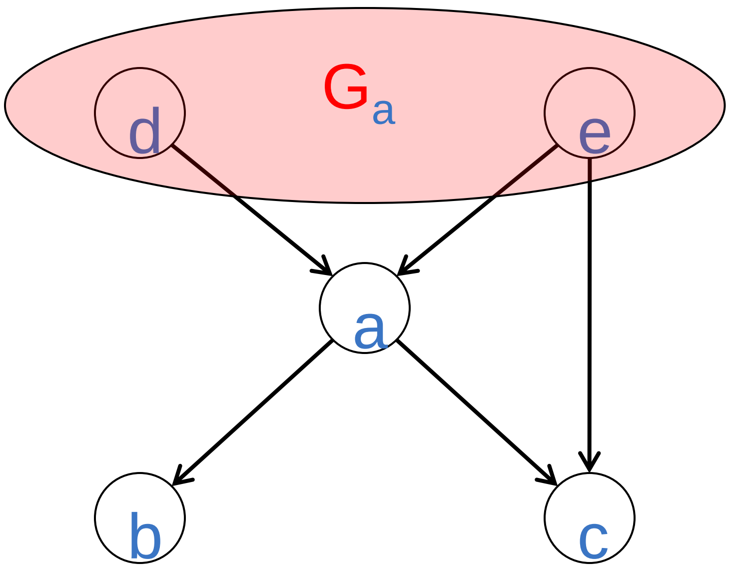
SoPU – Mission

Build and maintain a bridge that connects
algorithm theory and computational statistics
by developing the methodology of
computing large sums of products.


Victoria Falls Bridge. (Construction)

Probabilistic Models – Sums of Products

Bayesian network



Computational tasks

Inference:

$p_G(a|bc)$

$= \sum_{de} p_G(abcde) / \sum_{ade} p_G(abcde)$

Learning:

$G^* \in \mathrm{argmax}_G \; p_G(abcde)p(G)$,

with $p(G)=p(G_a)p(G_b)\ldots p(G_e)$

$p_G(abcde)=p(d)p(e)p(a|de)p(b|a)p(c|ae)$

Sums of Products – Algebra & Combinat.

Algebra

$$\sum_{x \in A} \prod_{S} f_S(x_S)$$

- Rings

  (+, ·) over integers

  (+, ·) over polynomials
- Semirings

  (max, ·)

  (min, +)

  (min, max)

Combinatorics

- The scopes $S \subseteq \{1,...,n\}$ form a hypergraph.

  (E.g., in BN inference)

- The summation is over a domain $A \subseteq D_1 \times \cdots \times D_n$ that may have a combinatorial structure.

  (E.g., in BN learning)

ALGODAN Algorithmic Data Analysis

ALGODAN Algorithmic Data Analysis


© Jamey Stillings/Barcroft Media

SoPU: Results 2010–2011

- Algorithm theory

  SODA'10,

  ICALP'10,

  SODA'12,

  Information Processing Letters 2010.

- Computational statistics

  AISTATS'10,

  UAI'11,

  ECML-PKDD'11,

  SDM'11.

Interactions: a directed acyclic graph

2012	SODA						
2011				UAI	ECML-PKDD		SDM’11
2010	IPL	ICALP	SODA			AISTATS	
2009	IWPEC		ESA		UAI		
2008	FOCS	STACS		ICALP		WABI	
2007			STOC				
2006	FOCS	UAI	COLT	IPL	ECML		
2005		ICML	JMLR	WABI			
2004		Thesis					
2003	PSB						

Permanent stuff (BHKK, IPL 2010)

- per $A = \sum_p a_{1p(1)} \cdots a_{mp(m)}$,

  where p runs through all injections from $[m]$ to $[n]$.

- Theorem

Algebraic structure	Time complexity
semiring	$m\, B(n, m)$
commutative semiring	$m(n-m+1)2^m$
ring	$m\, B(n, m/2)$
commutative ring	$(mn-m^2+n)2^m$

$B(n, m)$ is the number of subsets of $[n]$ of size at most m.

ALGODAN Algorithmic Data Analysis

What Next

- Keep the main themes
  - Make use of subtraction (additive inverses)
  - Optimization via counting
  - Space-time tradeoff considerations
- Bilinear transforms
  - Systematic study
- Bayesian networks
  - Implement into a public software
  - Apply to causal discovery with domain experts
- Other
  - Can randomized algorithms be much faster?
  - Better combinatorial bounds?