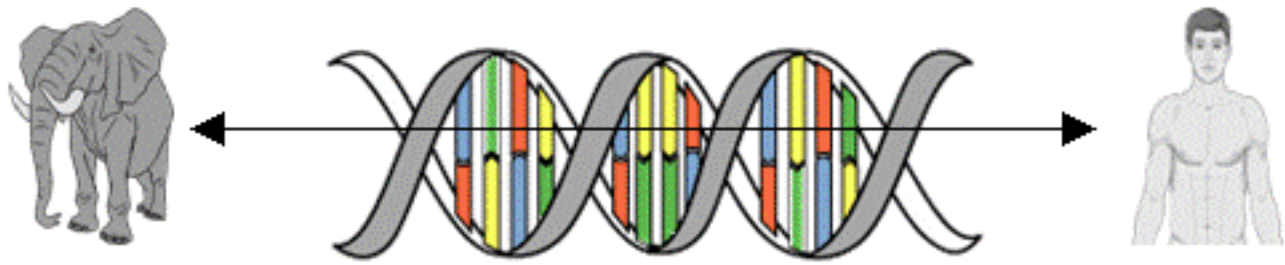
Aggregating Discrete Information from Mutually Inconsistent Sources

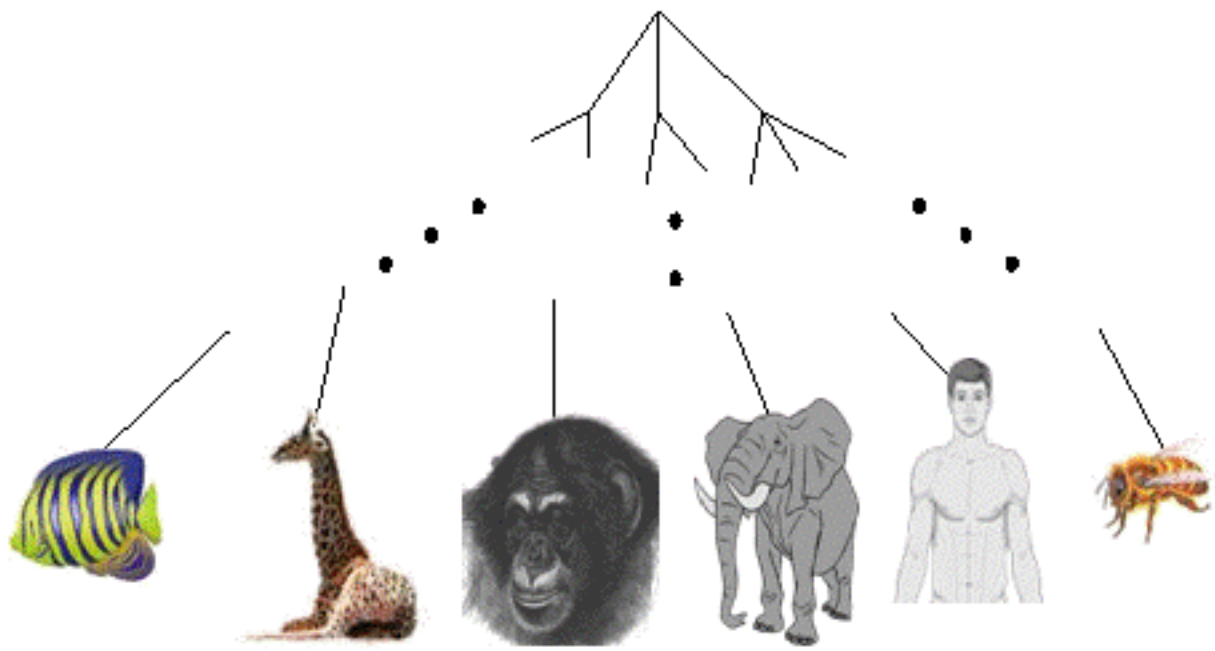
Nir Ailon
based on joint work with Moses Charikar and Alantha Newman

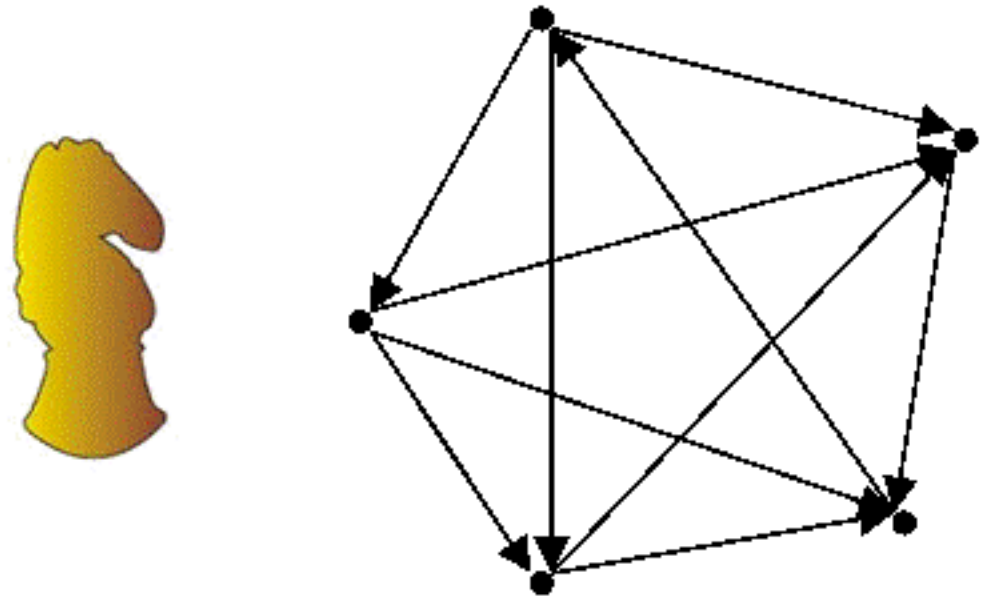
reconstructing . . . .

reconstructing . . . .

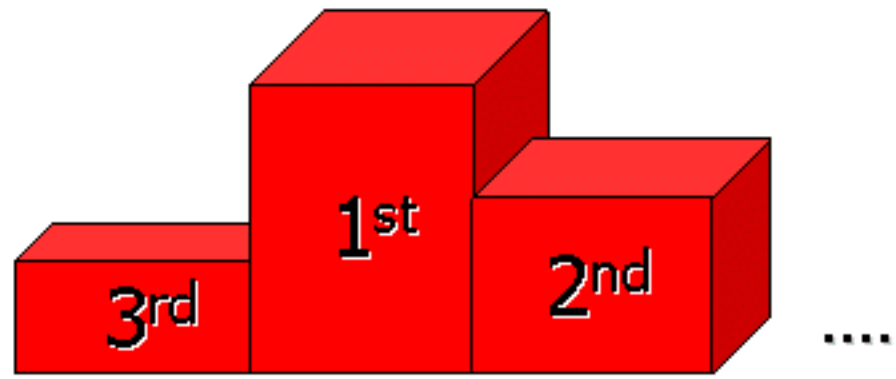


reconstructing . . . .

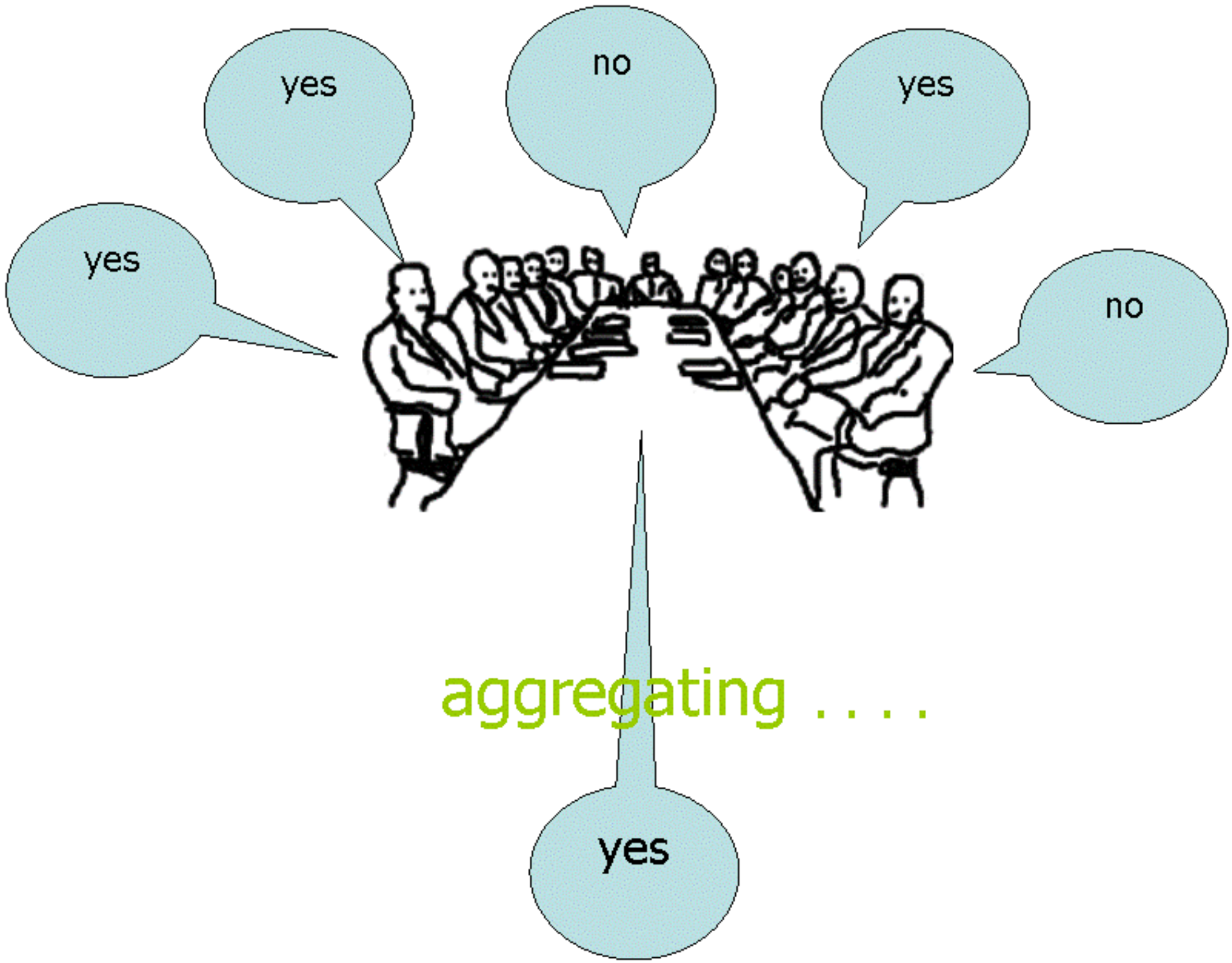


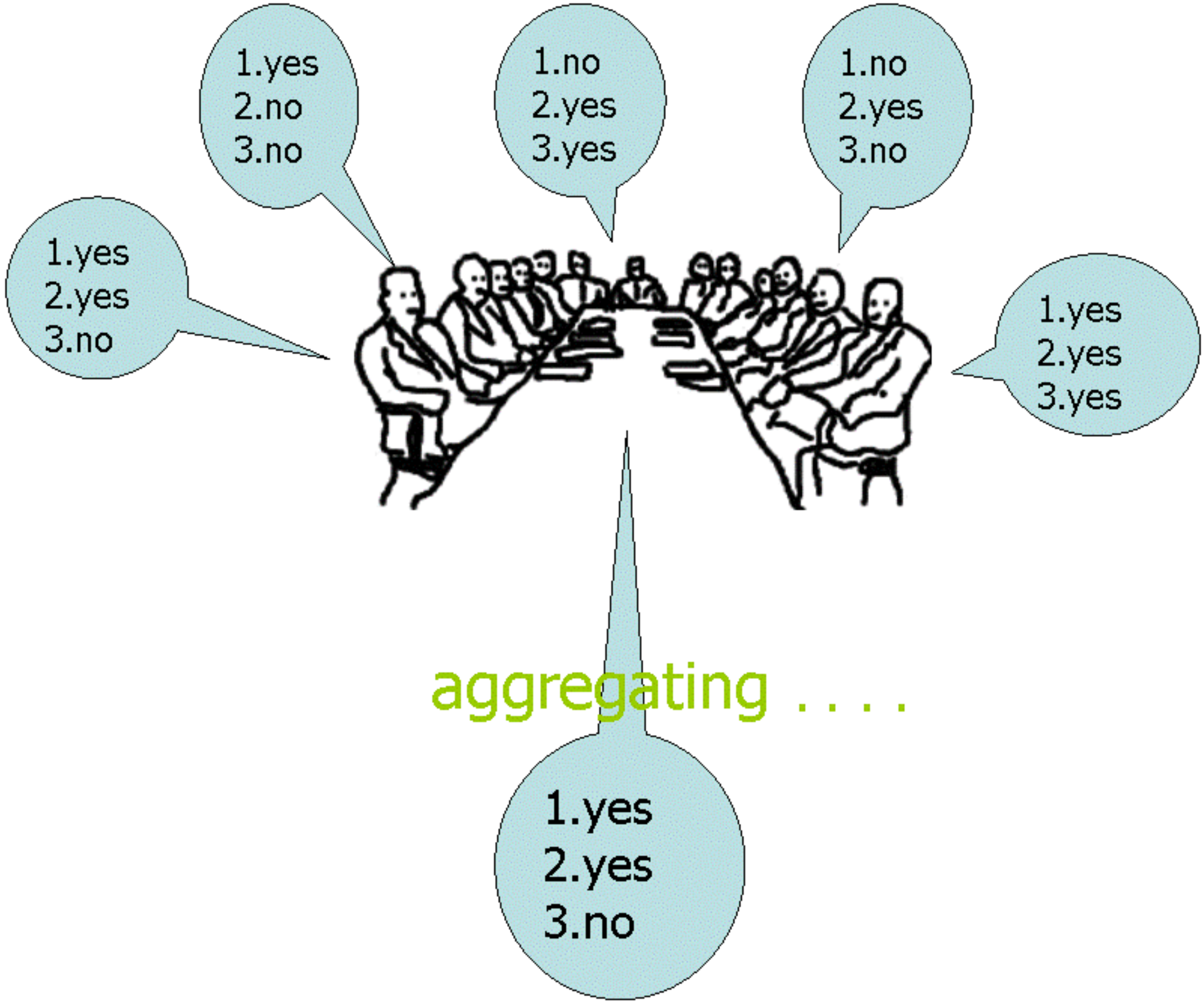


reconstructing . . . .

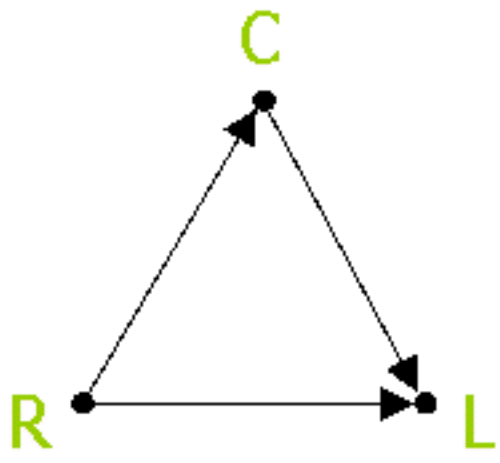


....

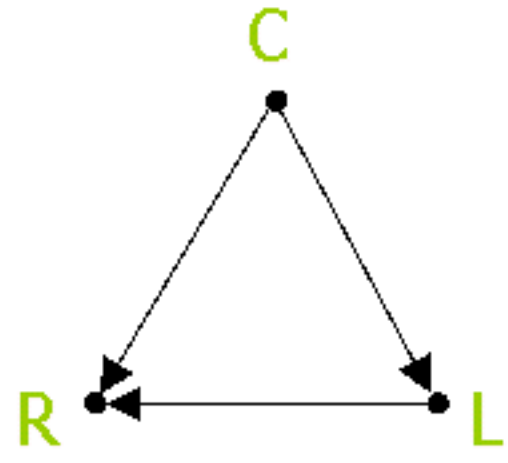

yes
no
yes
yes
no
aggregating . . . .
yes


1.yes
2.no
3.no
1.no
2.yes
3.yes
1.no
2.yes
3.no
1.yes
2.yes
3.no
1.yes
2.yes
3.yes
aggregating . . . .
1.yes
2.yes
3.no

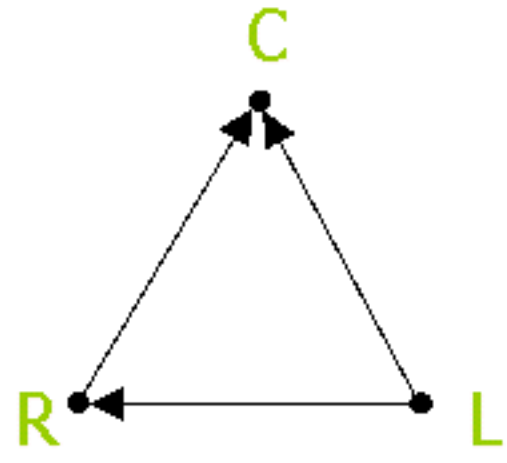
1. product Release before advertising Campaign?
2. advertising Campaign before massive Layoffs?
3. massive Layoffs before product Release?



R,C,L

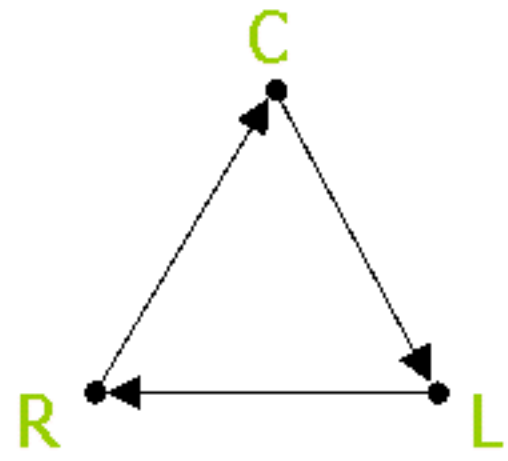


C,L,R



L,R,C

aggregating . . . .



???

Ranking Applications

- Information retrieval
- Voting (sports, politics...)
- Paleontology
- Finance
- Movie/hotel/restaurant rating
- Experimental psychology
- Clustering
- Collaborative filtering

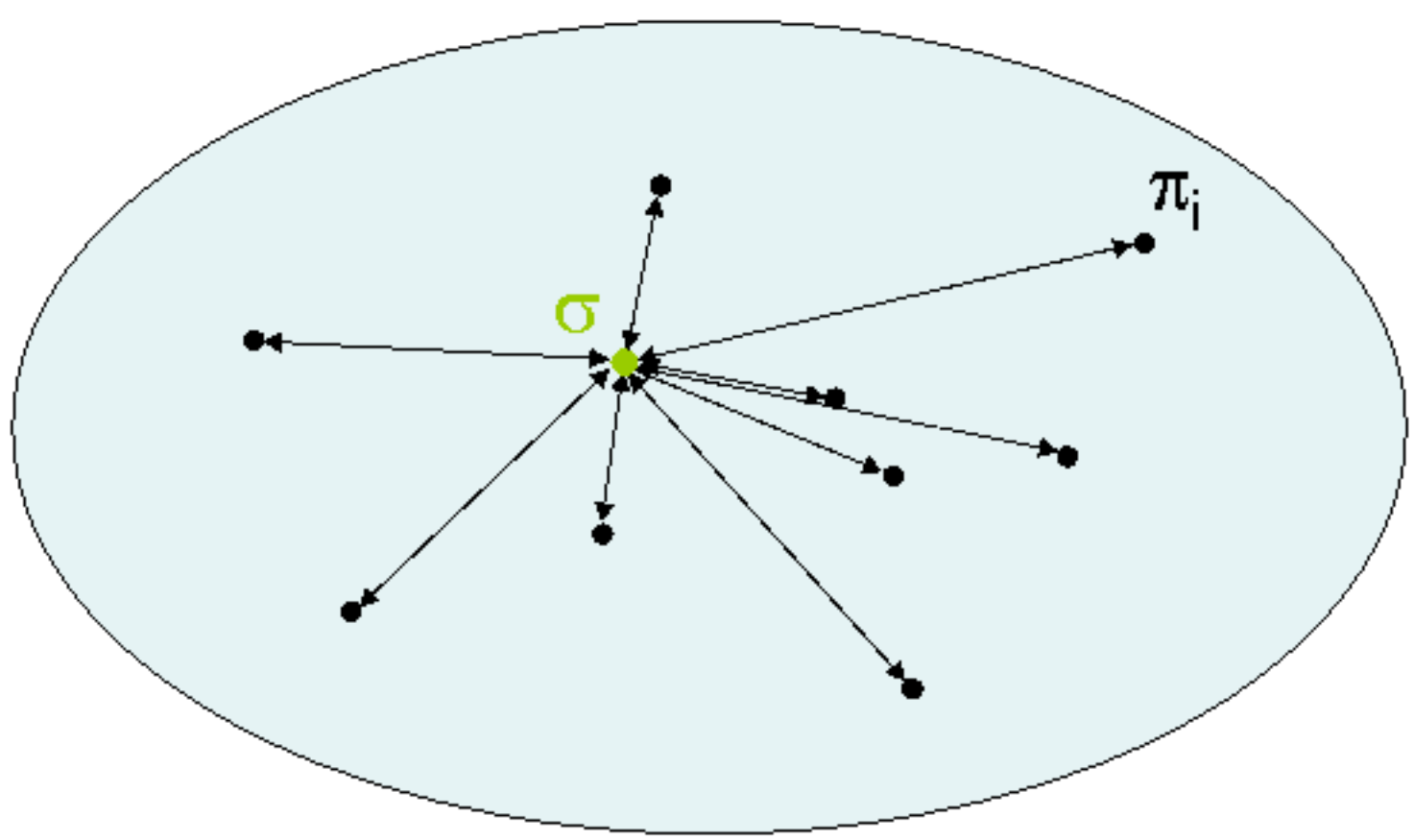
Rank Aggregation: *(Kemeny-Young, minimization)*

$$OPT = \min_\sigma cost(\sigma) = \min_\sigma \sum_{i=1..k} <\sigma, \pi_i>/k$$

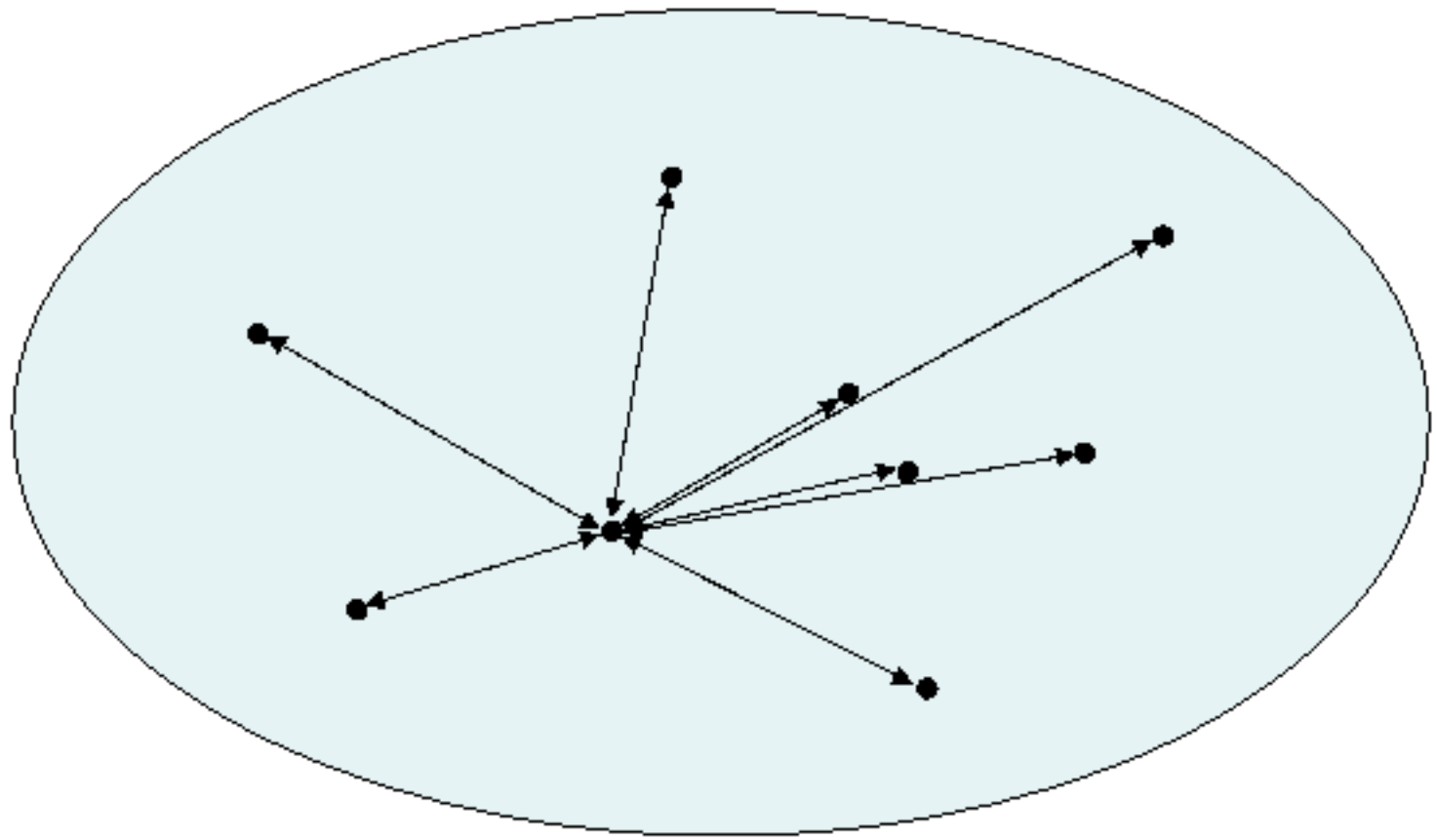
$<\sigma, \pi>$ = inversion distance

e.g. <ABC, CAB> = 2

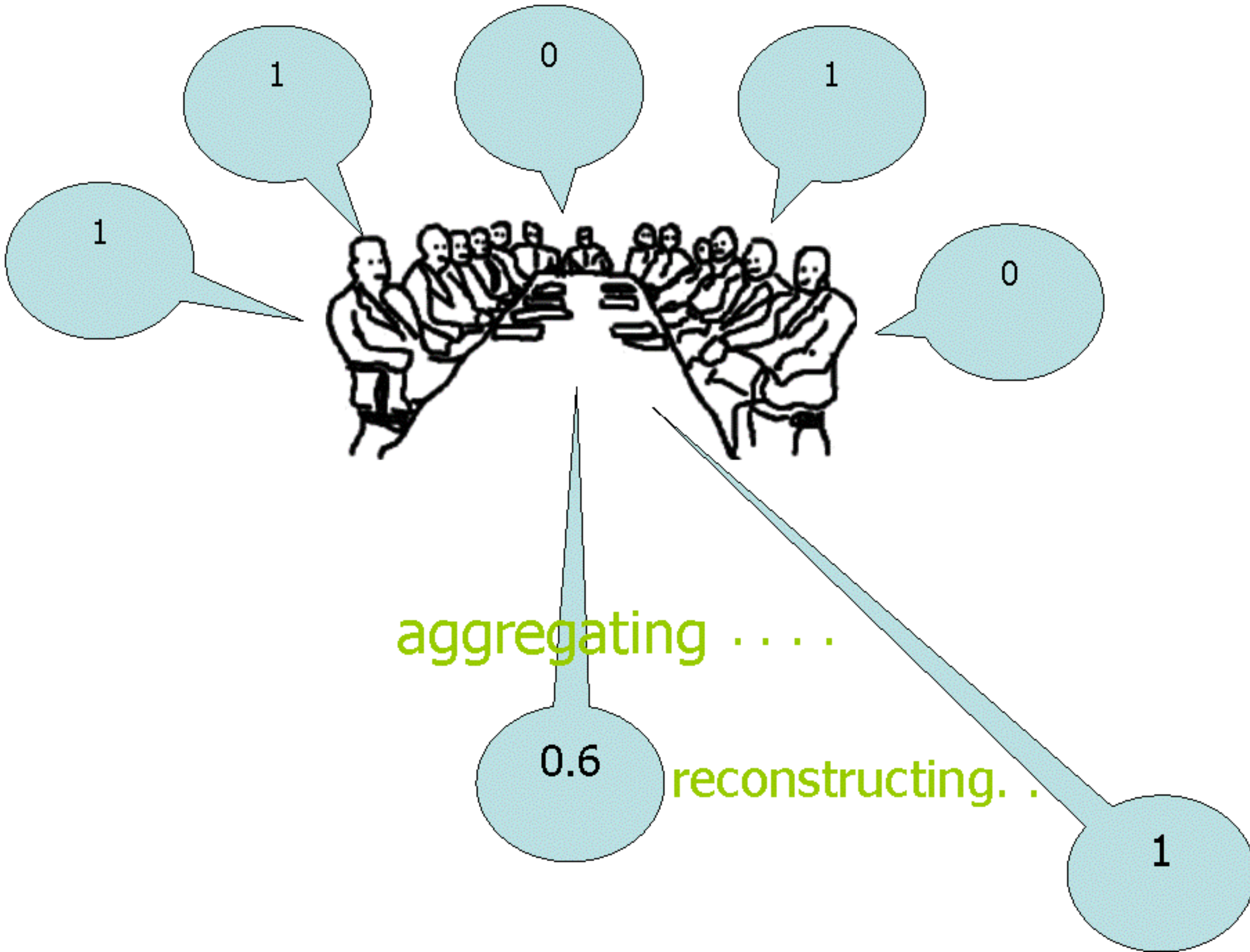
$<.,.>$ metric on permutations



$$\min_j \text{cost}(\pi_j) \leq 2\ \text{OPT}$$
$$E_{j \in [k]}\ [\text{cost}(\pi_j)] \leq 2\ \text{OPT}$$



Algorithm RandomChoice: 2-approx


1
0
1
1
0
aggregating . . . .
0.6
reconstructing. .
1

How to Take Average of Permutations?

write π as vector

score based: n coordinates

$\pi(u)$ = rank of $u \in V$

pairwise based: n^2 coordinates

$\pi(u,v) = 1$ if $u <_\pi v$, 0 otherwise

$<\pi,\sigma> = \sum \pi(u,v)\sigma(v,u)$

$\hat{\pi} = \sum_{i=1..k} \pi_i/k$

$\hat{\pi}(u,v) = \#\{i: u <_{\pi_i} v\}/k$

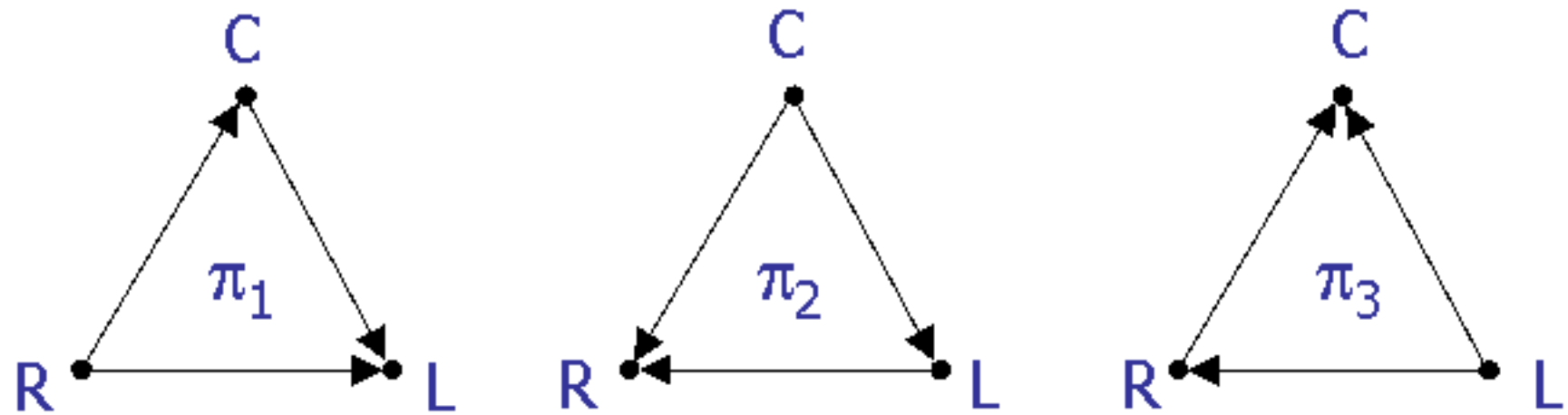
$\mathbf{cost}(\sigma) = <\sigma, \hat{\pi}>$

$= \sum_{u <_\sigma v} \hat{\pi}(v,u)$

$\hat{\pi} = \sum_{i=1..k} \pi_i/k$

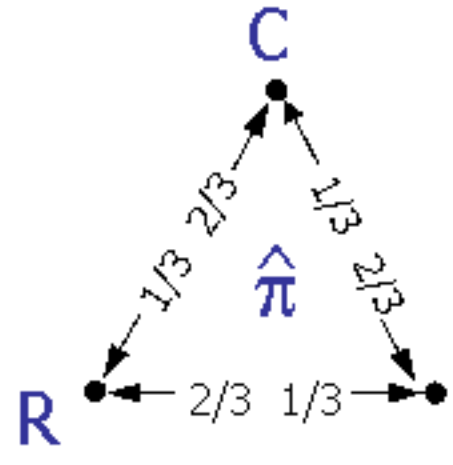
$\hat{\pi}(u,v) = \#\{i: u <_{\pi_i} v\}/k$

$\text{cost}(\sigma) = \langle \sigma, \hat{\pi} \rangle$



R,C,L C,L,R L,R,C

1. aggregating . . .



2. reconstructing. . . .

Minimum Feedback Arc-Set (min-FAS):

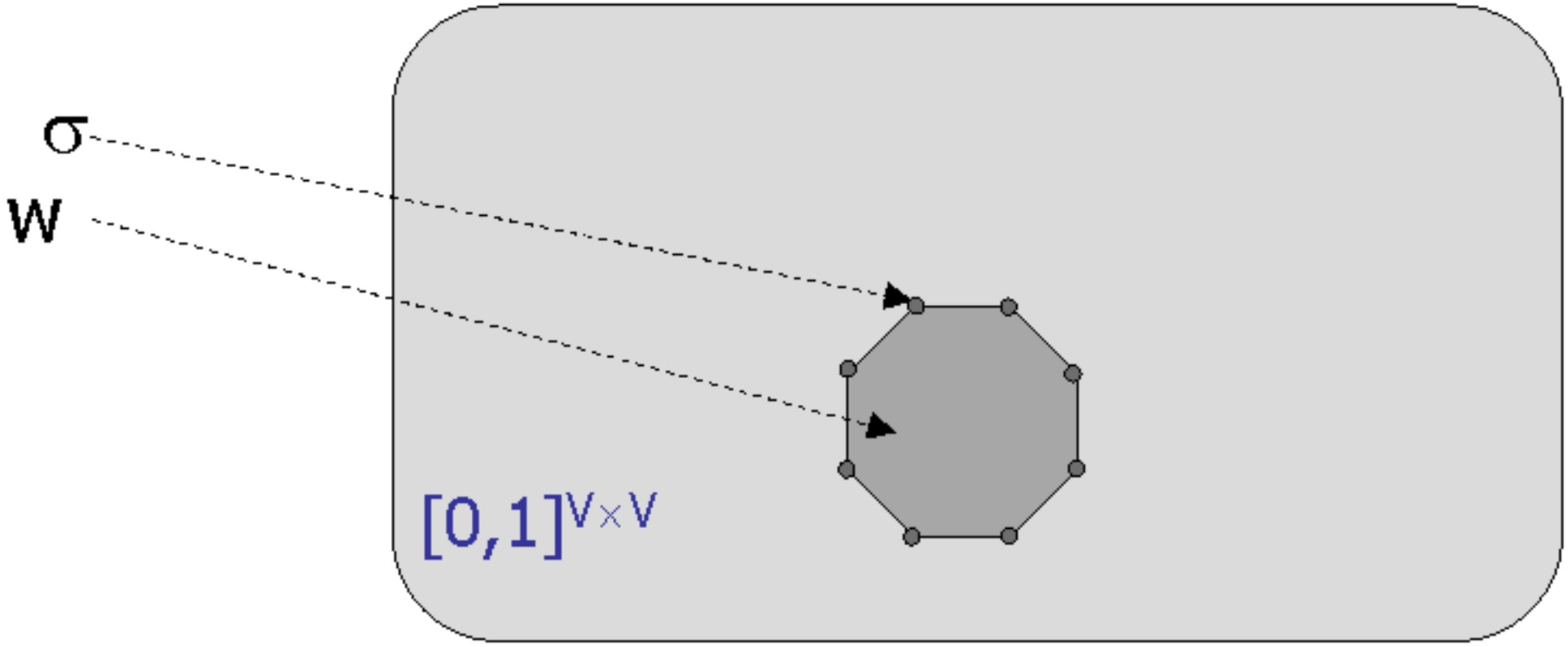
given:

set V of (n vertices)

$w: V \times V \rightarrow R^+$

minimize $cost(\sigma)=<\sigma, w>$ over permutations σ

Rank Aggregation: $w=\hat{\pi} \in$ conv(permutations)



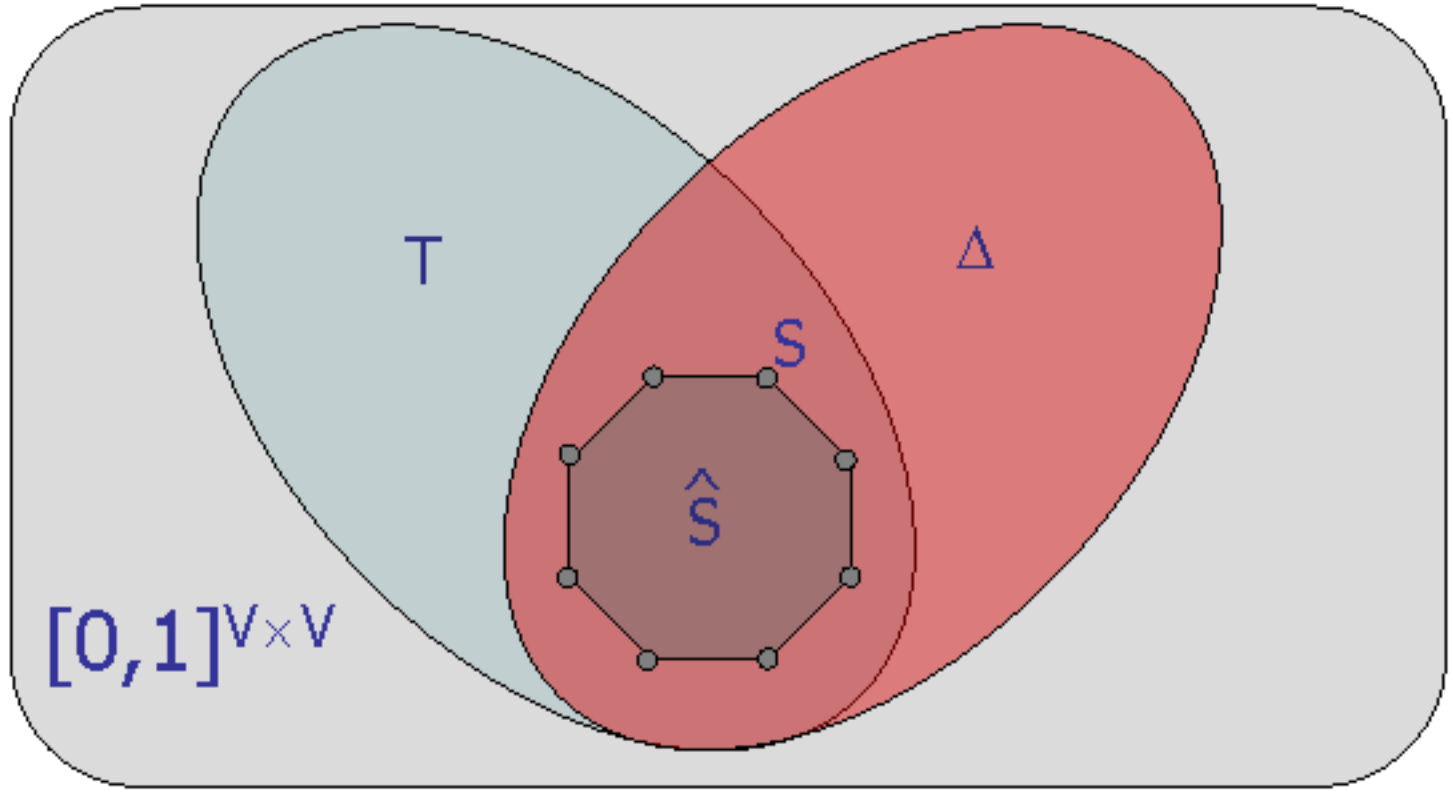
S: permutations on V

$\hat{S}$: convex closure of S

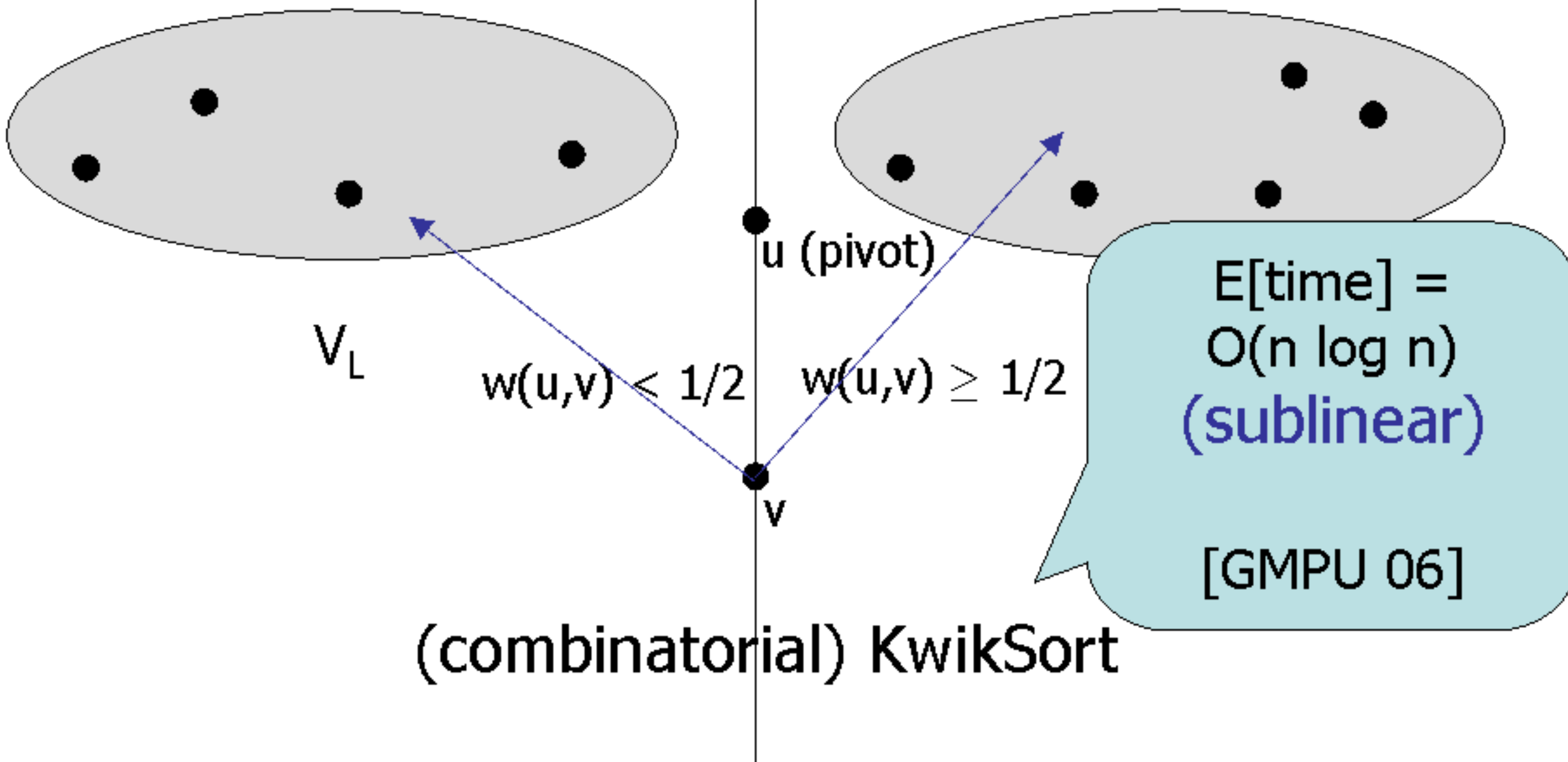
T: $\{w: w(u,v) + w(v,u) = 1 \ \forall \ u,v\}$

Δ: $\{w: w(u,v) \leq w(u,y) + w(y,v) \ \forall \ u,v,y\}$

$S = T \cap \Delta \cap \{0,1\}^{V \times V}$



KwikSort(V) = KwikSort(V_L), u, KwikSort(V_R)



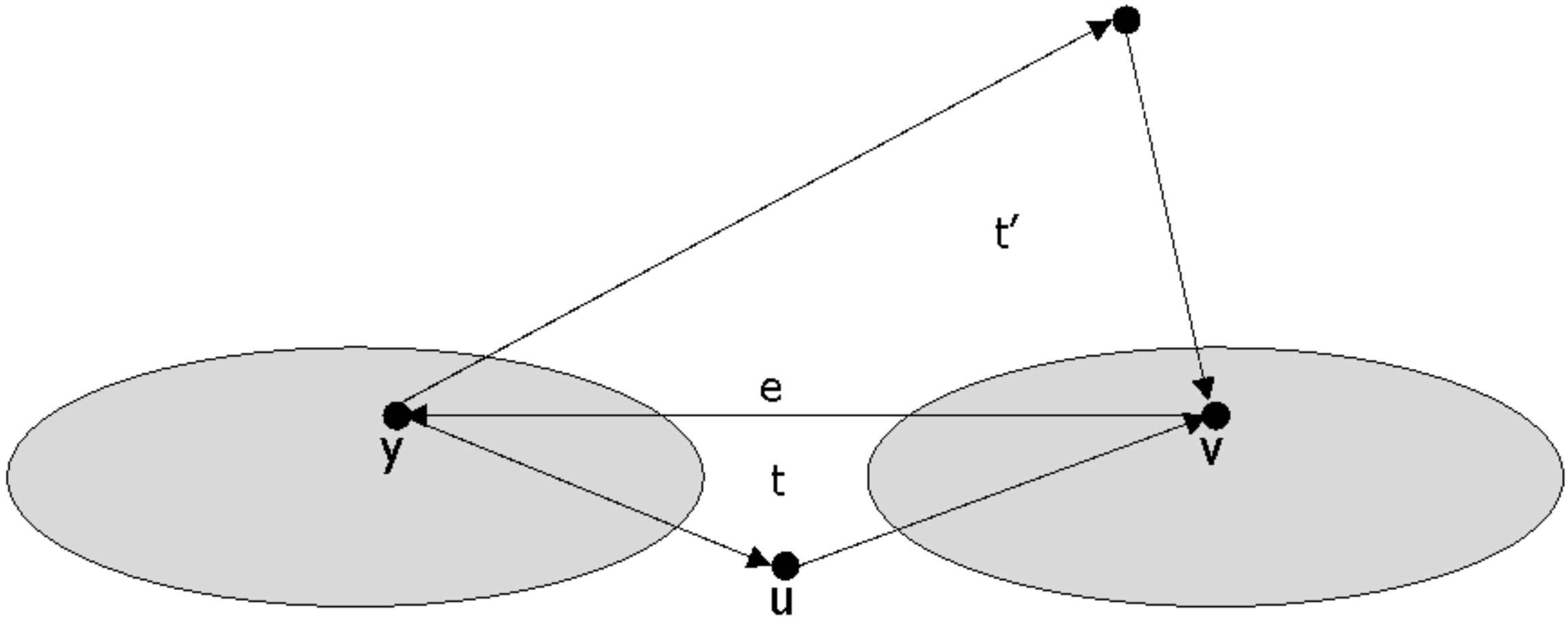
(combinatorial) KwikSort

(Combinatorial) KwikSort

input w	approx	previous	hardness
$[0,1]^{V\times V}$	?	log n loglog n	Max-SNP-Hard
T	5	log n loglog n	NP-Hard
$T\cap\{0,1\}^{V\times V}$	3 (t'nament)	log n loglog n	NP-Hard
Δ	2	log n loglog n	NP-Hard
Rank Agg	2	2	NP-Hard

best of two: 11/7

$T \cap \{0,1\}^{V \times V}$ 3 (t'nament) Proof:



A_t: "triangle t charged" B_e="edge e flipped"

$E[\text{cost}] = \sum_t \Pr[A_t]$

Edge e can be charged to only one triangle t

$\Rightarrow (A_t \cap B_e) \cap (A_{t'} \cap B_e) = \varnothing$, but $\Pr[A_t \cap B_e] = \Pr[A_t]/3$

$\Rightarrow \forall e \sum_{t:\, e \in t} \Pr[A_t]/3 \leq 1$

$\Rightarrow \sum_t \Pr[A_t]/3 \leq OPT$ (LP duality – see next slide)

A_t: "triangle t charged" B_e="edge e flipped"

E[cost] = $\sum t$ Pr[A_t]

Edge e can be charged to only one triangle t

$\Rightarrow (A_t \cap Be) \cap (A_{t'} \cap B_e) = \varnothing$, but Pr[$A_t \cap B_e$]=Pr[$A_t$]/3

$\Rightarrow \forall e \sum t: e \in t$ Pr[A_t]/3 ≤ 1

$\Rightarrow \sum t$ Pr[A_t]/3 $\leq$ OPT (LP duality)

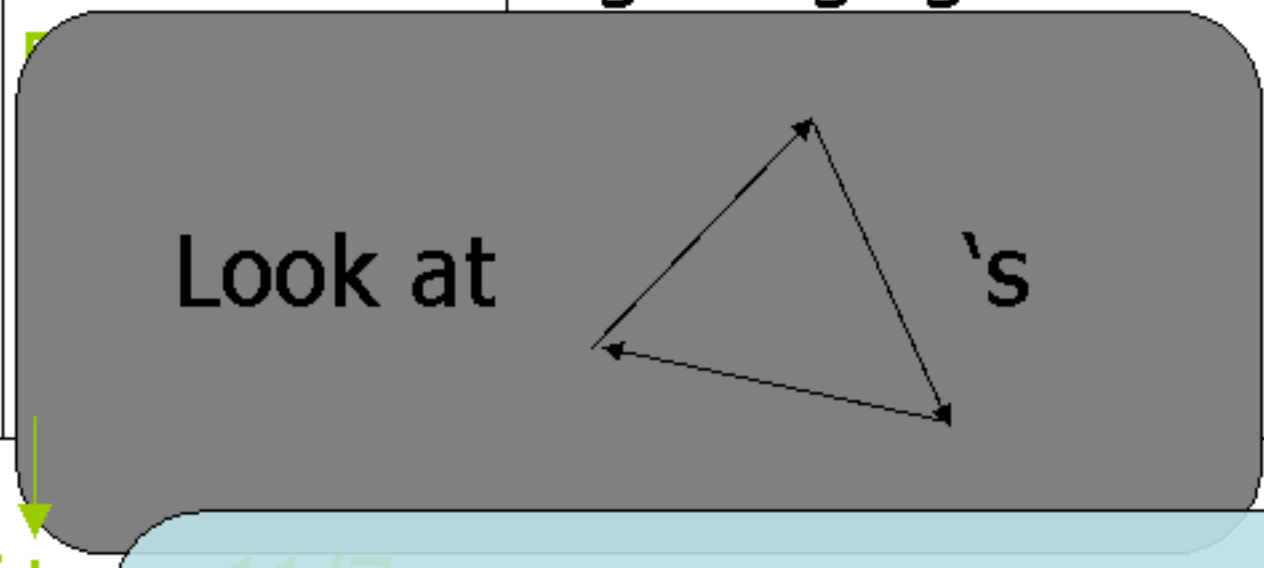
Primal: (hitting LP)
minimize total value of weights assigned to edges such that weights incident to each directed triangle sum up to at least 1

Dual: (packing LP)
maximize total value of weights assigned to directed triangles such that weights incident to each edge sum up to at most 1

How to Prove Other Approx Factors?

(Combinatorial) KwikSort

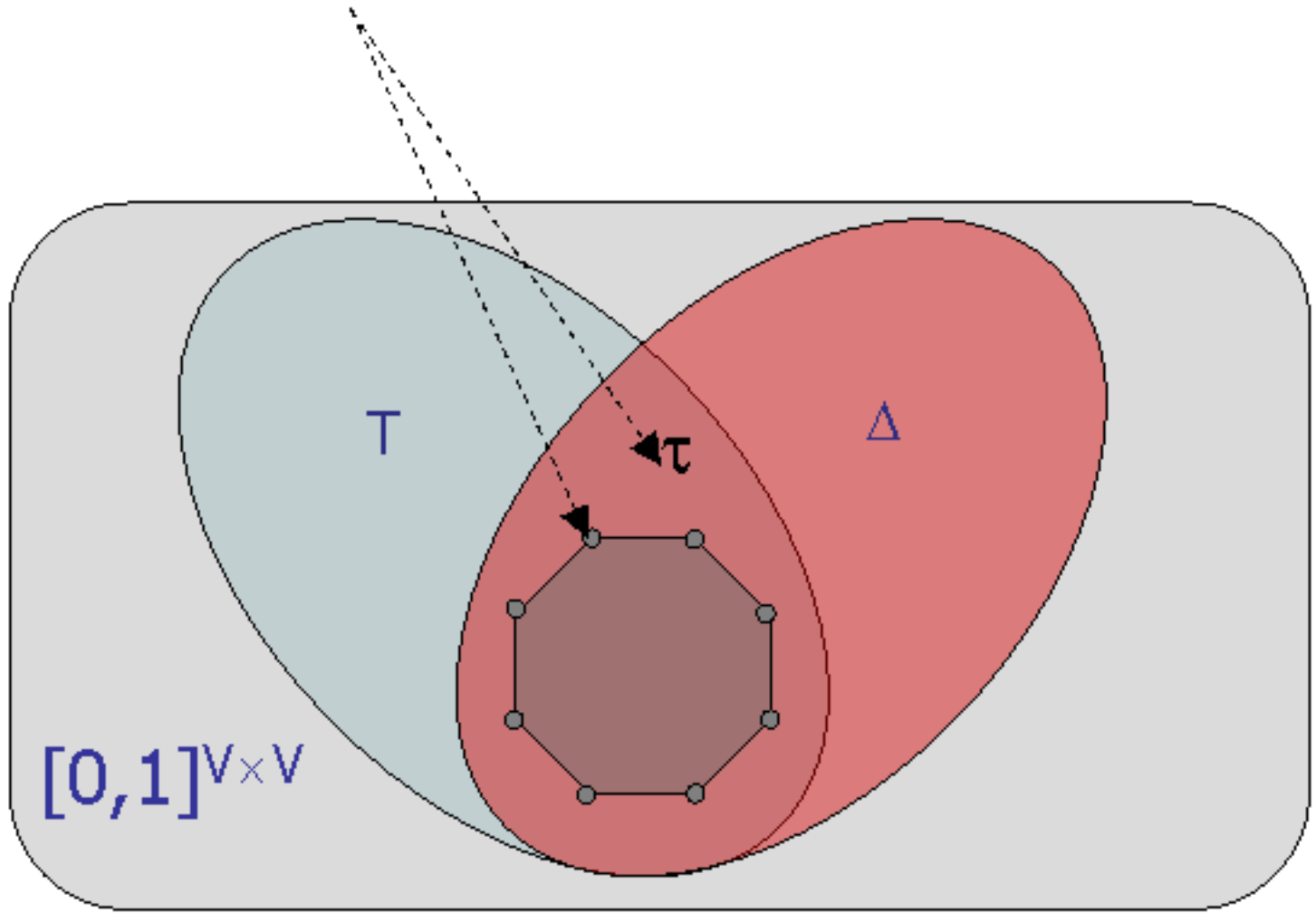
input w	approx	previous	hardness
$[0,1]^{V\times V}$	?	log n loglog n	Max-SNP-Hard
T			NP-Hard
$T\cap\{0,1\}^{V\times V}$			NP-Hard
Δ			NP-Hard
Rank Agg			NP-Hard



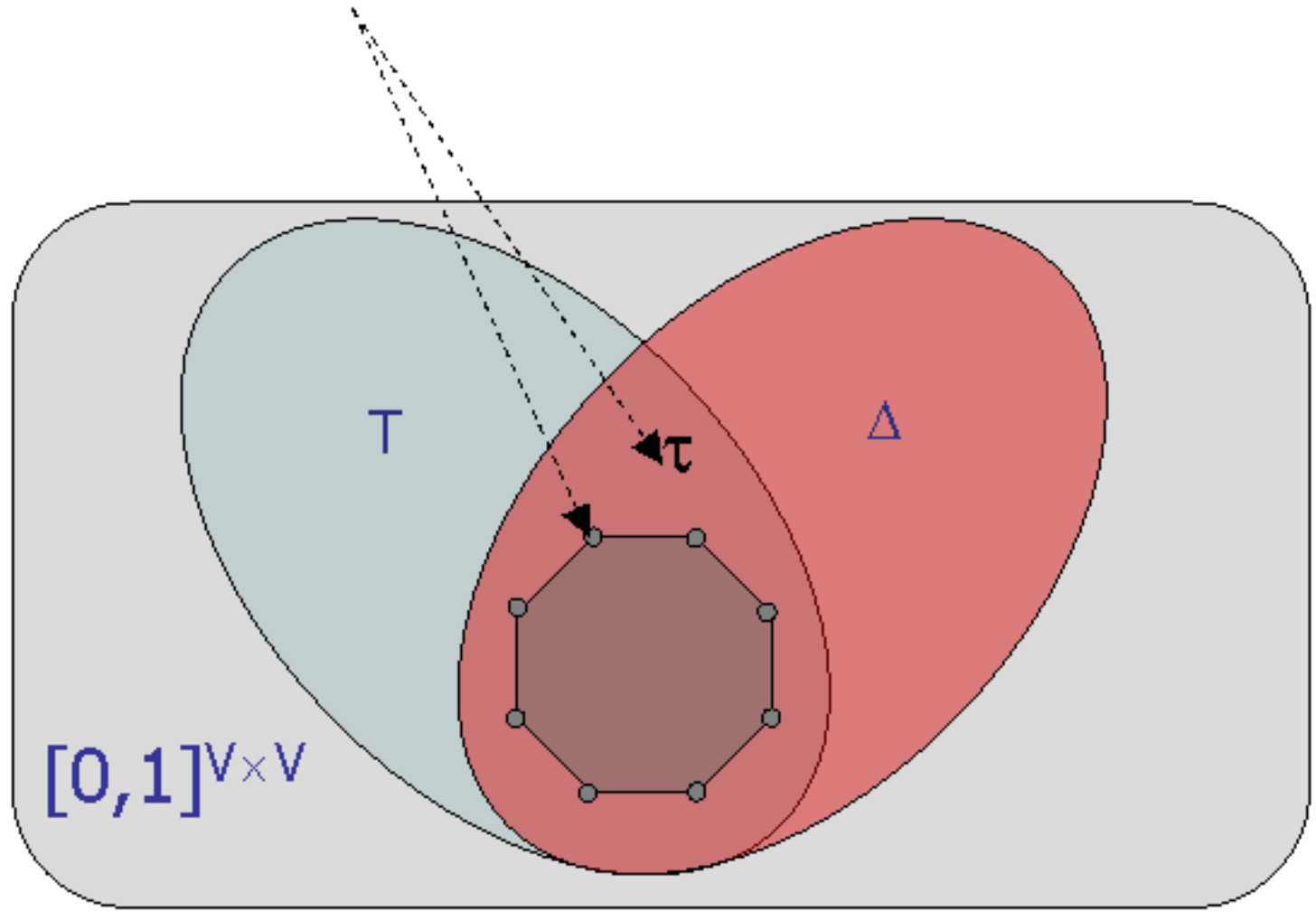
best of two: 11/7

Any structural/statistical information
on weights around triangles
in the input could translate
to improved approximation

optimize over $T \cap \Delta \cap \{0,1\}^{V \times V}$ (IP)



optimize over $T \cap \Delta$ (LP)



u (pivot)

with probability $\tau(v,u)$

with probability $\tau(u,v)$

v

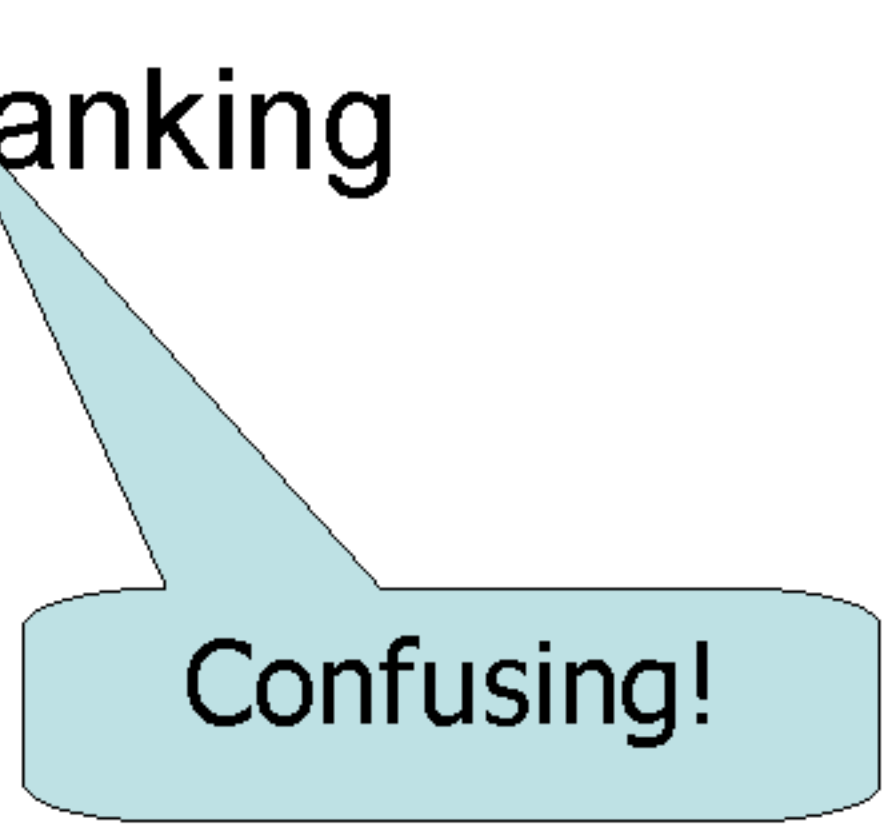
(LP-based) KwikSort

(LP-based) KwikSort

input w	approx	previous	hardness
$[0,1]^{V \times V}$	?	log n loglog n	Max-SNP-Hard
T	2.5	5	NP-Hard
$T \cap \{0,1\}^{V \times V}$	2.5(t'nament)	3	NP-Hard
$\Delta \cap T$	2	2	NP-Hard
Rank Agg	2	2	NP-Hard

best of two: 4/3 11/7

Ranking with Ties
a.k.a.
Bucket Order
a.k.a.
Partial Ranking



File Edit View Go Bookmarks Tools Help

http://www.library.uiuc.edu/edx/rankbus.htm

Go

Customize Links Free Hotmail file:///C:/Documents... RealPlayer Windows Marketplace Windows Media Windows BBC NEWS | Help | R...

General/Undergraduate

Graduate/Research Programs

Business

Law

International (Non-U.S.)

Humor & Satire

Bibliography

About This Site

Comments/Questions

Education & Social Science Library

UIUC Library Gateway

University of Illinois at Urbana-Champaign

Controversy page. For questions or comments, please contact Nancy O'Brien.

BUSINESS SCHOOLS

Asia's Best MBA Schools 2000

[http://www.asiaweek.com/asiaweek/features/mba/index.html]

Methodology: This site does not offers an explanation of its rankings.

(Please note: we have recently been informed that Asiaweek will no longer be publishing their annual rankings. We will continue to link to the site as long as it is relevant.)

Asiaweek offers its assessment of the top MBA programs in the Asia-Pacific region. The ratings are based on six categories: academic resources, student selectivity, faculty resources, university connections, grade output, and other resources such as library spending and technology. Separate tables are provided for the best full-time, part-time, executive, and distance programs as well as for graduate salaries.

B-School Net

[http://www.b-school-net.de/]

Methodology: This site offers an explanation of its rankings on the opening page.

This site, in German with English in places, offers recent rankings of Business Schools and leading MBA programs in Germany. There are several rankings including one based on a survey of 13,000 students focusing on several categories including: Quality of professors in teaching; Support for students by professors; Quality and actuality of library; Quality and availability of IT; Amount and quality of student activities; Usefulness in practice; Internationality; and Cooperation of school with corporations. There are also links to further information.

Business School Research Rankings

[http://www.bus.indiana.edu/ardennis/rankings/]

Methodology: This site offers an explanation of its rankings on the page titled Study Background.

Business School Research Rankings is intended to supplement a study of U.S. business school research performance published in the December issue of the *Academy of Management Journal*. The schools are

Done

$v_{300}, v_{250}, v_{1845}, \cdots, v_4, v_{50}, v_{1300}$

top-k aggregation (k=3) NP-Hard for $k \geq 2$

ties due to infeasibility

ranking of 2000 employees

$V=\{v_1, \cdots, v_{2000}\}$

Tools Help
Search Favorites
oo.com/marketupdate/newcoverages
Go Bookmarks 44 blocked Check AutoLink AutoFill Send to

Tuesday, May 29, 2007, 4:40PM ET - U.S. Markets Closed.

INVESTING

FEATURES

MARKET COVERAGE
Market Overview
Market Update
In Play®
Story Stocks
Short Stories
Tech Stocks

MARKET EVENTS
Earnings Calendar
Economic Calendar
Splits Calendar
Conference Call Calendar
IPO Calendar
Upgrades/Downgrades

MARKET STATISTICS
Market Digest
U.S. Market Indices
World Market Indices
Currency Exchange Rates

New Coverages

Upgrades | Downgrades

Company Name	Symbol	Analyst	Rating
Allergan	AGN	Lazard Capital	Buy
AMIS Holdings	AMIS	CIBC Wrld Mkts	Sector Outperform
Javelin Pharmaceutic	JAV	JP Morgan	Overweight
MasterCard	MA	Bear Stearns	Outperform
MetroPCS	PCS	UBS	Buy
MetroPCS	PCS	Bear Stearns	Outperform
MetroPCS	PCS	Banc of America Sec	Buy
MetroPCS	PCS	Wachovia	Outperform
Panacos Pharma	PANC	Punk, Ziegel & Co	Buy
Sirona Dental Systems	SIRO	Banc of America Sec	Neutral

Take a FREE TRIAL of Briefing.com's complete Live market analysis of the U.S. stock and bond markets, technology stocks, economic releases, earnings reports, and day trading highlights

Copyright © 2007 Yahoo! All rights reserved. Privacy Policy - Terms of Service - Copyright Policy

Review: New York - New York Hotel and Casino, Las Vegas, Nevada

1 2 3 4 5 **4.0** New York, New York, Timeless!!!!

April 03, 2006 *A TripAdvisor Member*, Minneapolis, MN

... to Vegas and loved **New York, New York**. Our room ... ... will always stay at **New York, New York**. Many of the other casinos seemed dated or were way too expensive, but **New York** is timeless ...

Review: New York City, New York

1 2 3 4 5 **5.0** I love New York

January 05, 2006 *A TripAdvisor Member*, Columbus,OH

Good place to visit, bad place to live.

Review: New York - New York Hotel and Casino, Las Vegas, Nevada

1 2 3 4 5 **1.0** New York New York Hotel

February 01, 2005 *A TripAdvisor Member*, Los Angeles

We stayed in **New York New York** hotel bcos ... ... is looking for staying in **New York New York** bcos they are offering 2 SPA (worth of 40$?????) you better change your mind

Review: Ramada Plaza & Inn - New Yorker Hotel, New York City, New York

1 2 3 4 5 **5.0** New York

March 26, 2003 *A TripAdvisor Member*

... is Madisson Square Garden and Times Square was 15 minutes walk away. I WILL be returning to **new york** and when I do I will be staying in the **New** Yorker again

See the next 20 results

minimize
$\langle \sigma, \hat{\pi} \rangle$
over $\sigma \in S$

p-rating aggregation
(p=3)
in P for $p \leq 2$

π_i

v_1 ☺ ☺
v_2 ☺ ☺ ☺
v_3 ☺
v_4 ☺ ☺
v_5 ☺ ☺
v_6 ☺ ☺ ☺
v_7 ☺
v_8 ☺
v_9 ☺ ☺ ☺
v_{10} ☺ ☺ ☺

$\pi_i \notin \mathsf{T}$
$\pi_i \in \Delta$

ties due to coarseness

ranking of 10 executives
$V=\{v_1,\ldots, v_{10}\}$

$\pi_i(v_{10},v_1)=1$
$\pi_i(v_1,v_{10})=0$
$\pi_i(v_5,v_4)=\pi_i(v_4,v_5)=0$

$v_{10} < v_1$
$v_5 \equiv v_4$

minimize
$\langle \sigma, \hat{\pi} \rangle$
over $\sigma \in S$

e.g. <ABC, [CA]B> = 1

If some voter ties u,v together
no price incurred for u,v

Why not pay for ties?

Can pay for ties but "easier" for approximation

Why minimize over permutations?

Depends on "amount of information" available and appication

<·,·>
not metric on
$\Delta \cap \{0,1\}^{v \times v}$
how to get 2
approx?

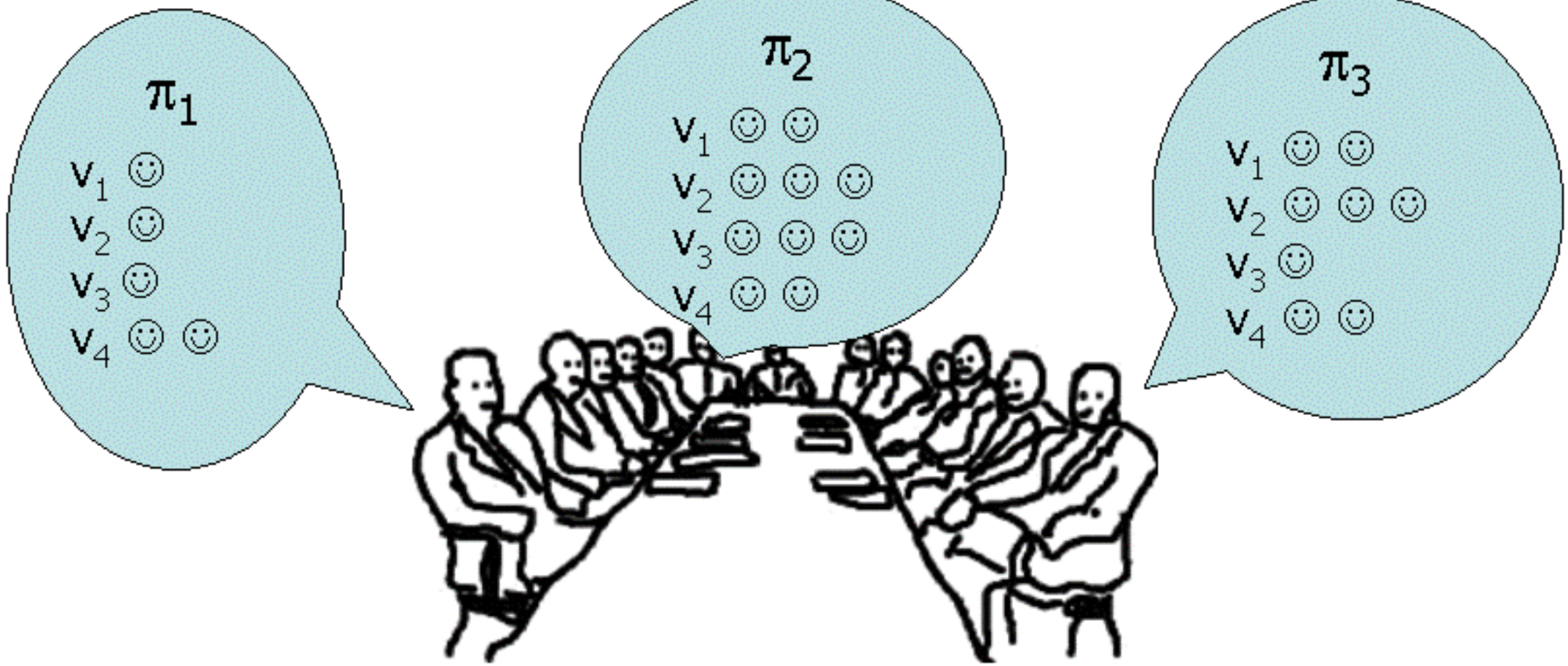
$P_i^{-1}(\pi_i)$

π_i' (permutation)

$\langle \sigma, \pi_i \rangle$

σ (permutation)

$\pi_i = P_i(\pi_i') \in \Delta \cap \{0,1\}^{v \times v}$



ranking of 4 executives

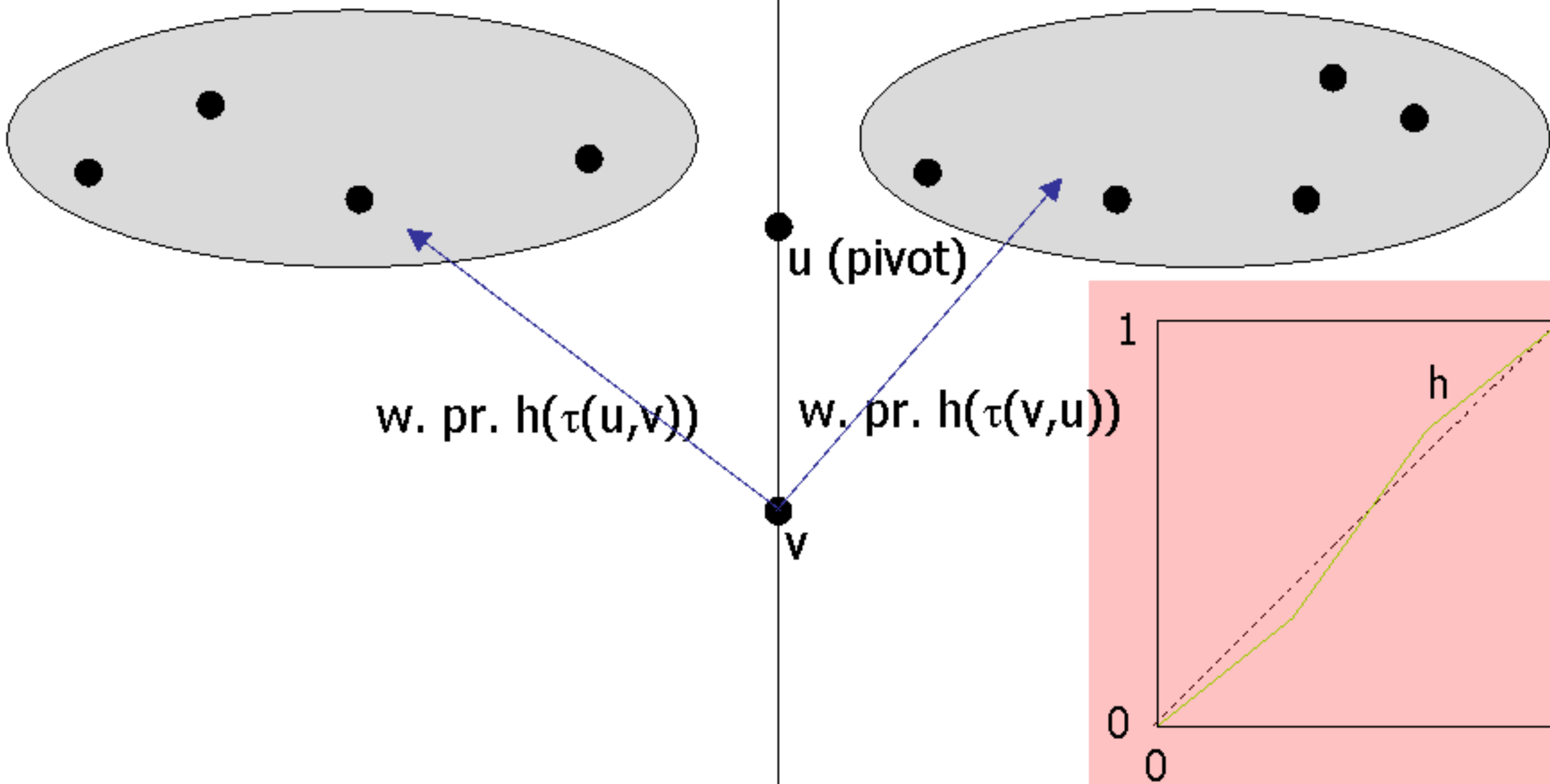
$V=\{v_1, v_2, v_3, v_4\}$

algorithm RepeatChoice: 2 approximation

$[v_2\ v_3]$, $[v_1\ v_4]$ (chose voter #2)
v_2, v_3, $[v_1\ v_4]$ (broke ties with voter #3)
v_2, v_3, v_4, v_1 (broke ties with voter #1)

3/2 approximation algorithm:

conjecture:
best of RepeatChoice
and this gives
4/3 approximation

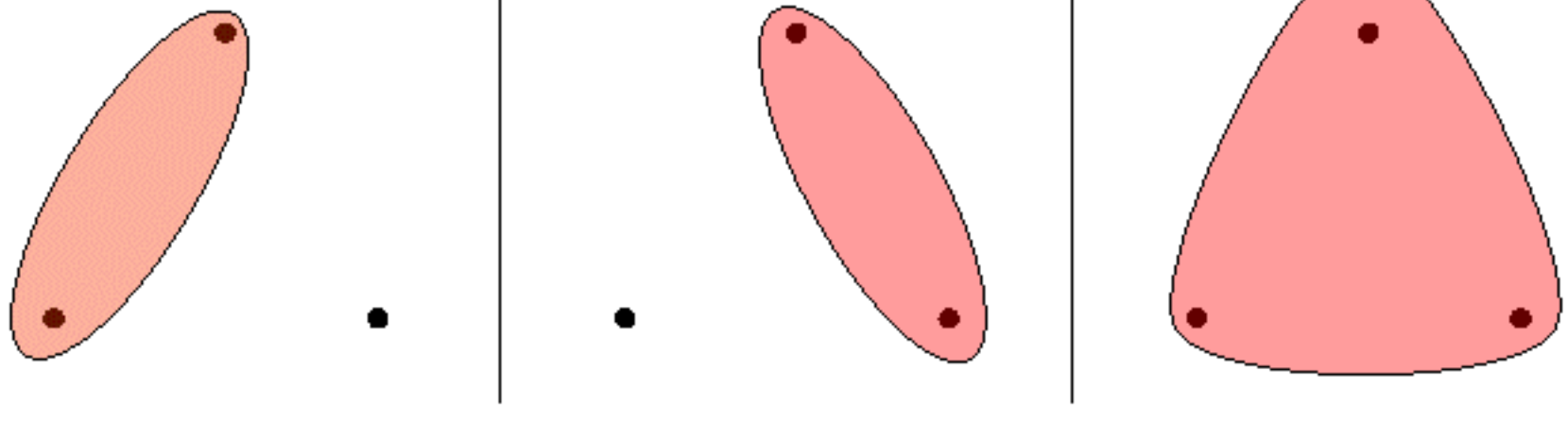


Followup Work

- Coppersmith, Fleischer, Rudra:
  Borda meets Condorcet
- Hedge, Jain, Williamson, Van Zuylen:
  Derandomization
- Mathieu-Kenyon, Schudy:
  PTAS!

ranking with ties = ranking + clustering

clustering



aggregating . . . .

Correlation Clustering (on complete graphs):

given set V of n vertices, $w: \binom{V}{2} \to R^+$

minimize $\langle \sigma, w \rangle$ over clustering σ

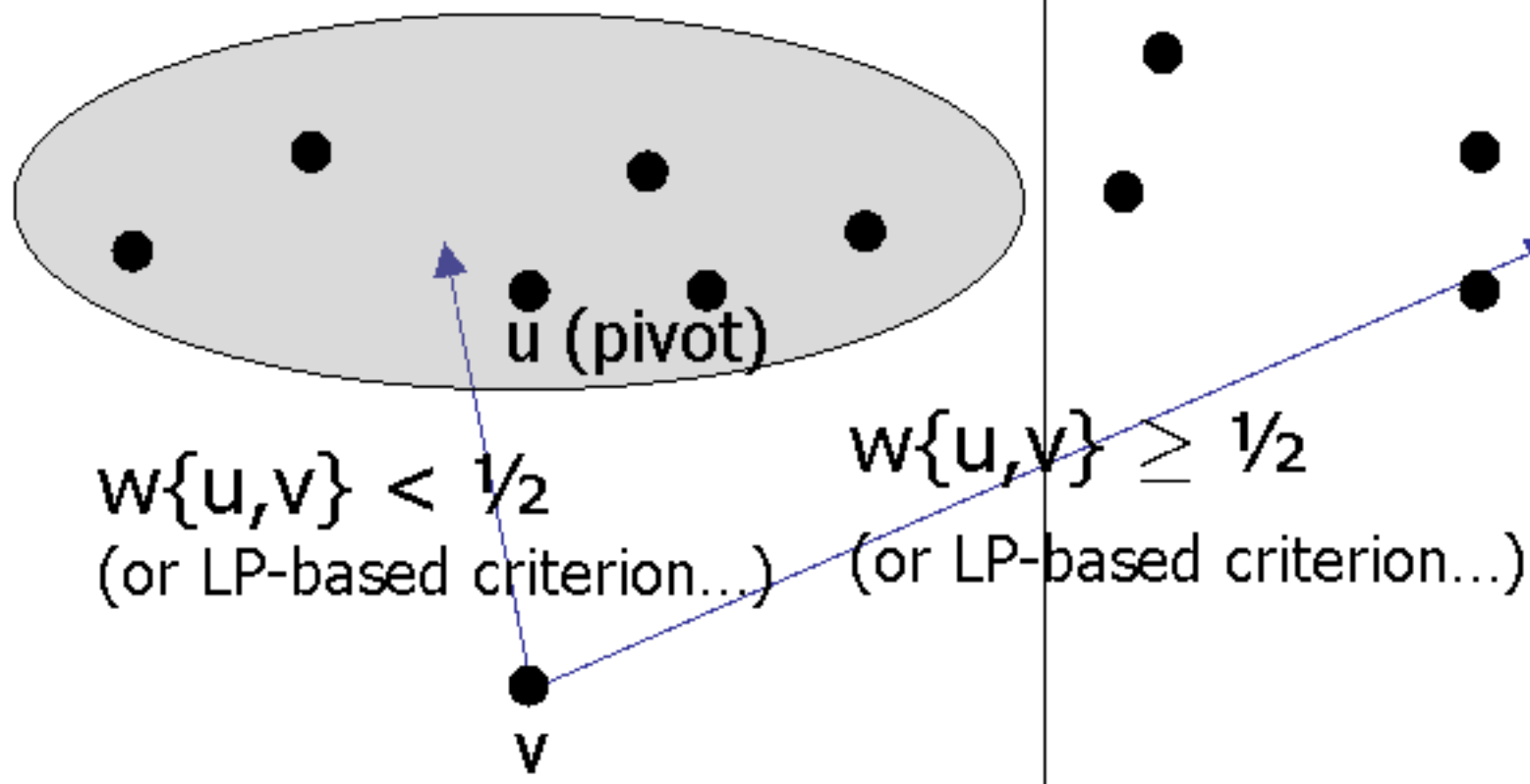
pay $w\{u,v\}$ whenever $u \equiv_\sigma v$

pay $1-w\{u,v\}$ whenever $u \not\equiv_\sigma v$

Consensus Clustering:

$w \in \text{conv}(\text{clusterings})$

Time can be quadratic w. prob. 1 for some inputs

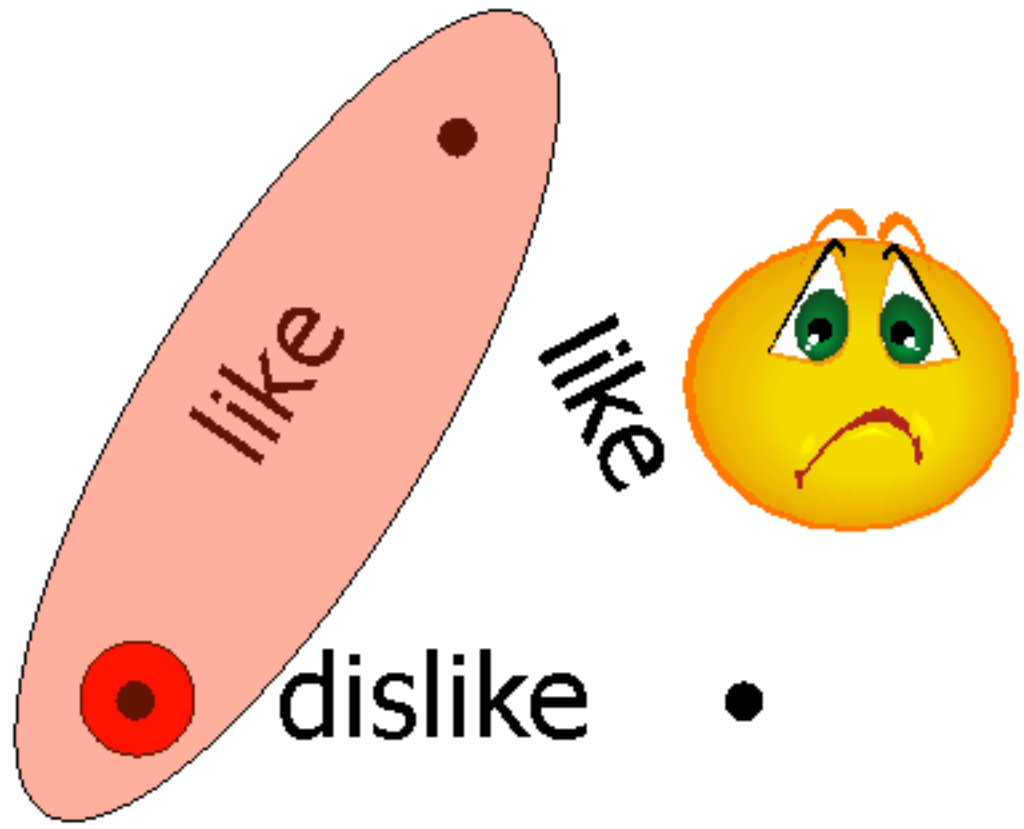


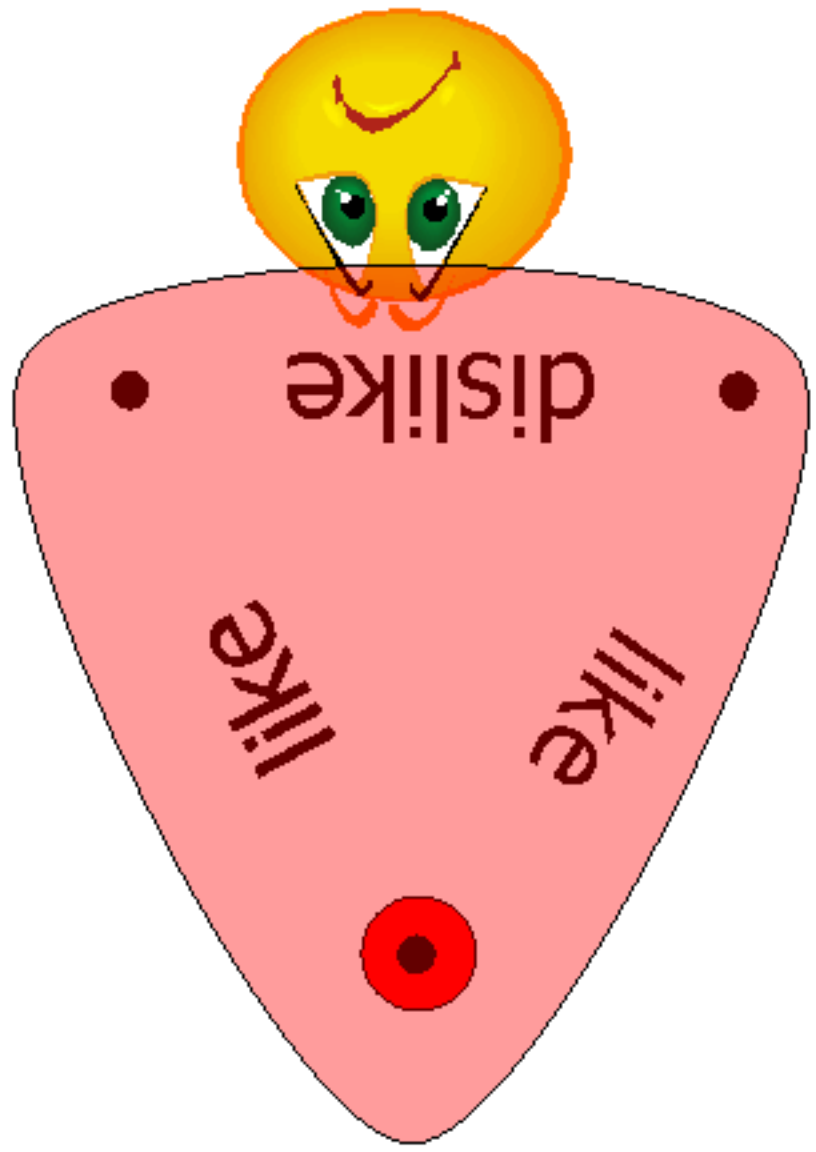
(Combinatorial/LP-based) KwikCluster

"LP-based" KwikCluster

input	approx	previous	hardness
$[0,1]^{V \times V}$	2.5(corr-clust)	4	Max-SNP-Hard
$\{0,1\}^{V \times V}$	2.5	4	Max-SNP-Hard
Δ	2	4	Max-SNP-Hard
Cons Clust	2	2	Max-SNP-Hard

best of two: 4/3


like
like
dislike


dislike
like
like