



**FLASHPOINTS:** MINING EXCEPTIONAL  
PAIRWISE BEHAVIOR  
IN VOTE OR RATING DATASETS

**AUTHORS.**

**Adnene BELFODIL** Philippe LAMARRE  
Marc PLANTEVIT Sylvie CAZALENS



Collaborative  
rating platforms



MovieLens



Yelp



Voting datasets



European  
parliament voting



Collaborative  
rating platforms



Movielens



Yelp



## Collaborative rating platforms



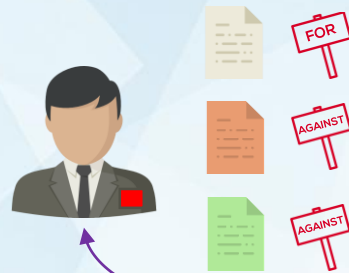
Age, Occupation



Category, launch year...



## Voting datasets



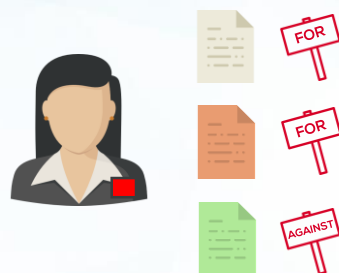
Political group, country...




Theme, ballot date ...

\*context

⋮  
other  
deputies



  
Voting datasets

 European  
parliament voting

Left wing

Right wing

Left wing female deputy: 4 FOR, 2 AGAINST

Left wing male deputy: 3 FOR, 2 AGAINST

Left wing female deputy: 4 FOR, 2 AGAINST

Left wing male deputy: 4 FOR, 2 AGAINST

other deputies

Right wing female deputy: 4 FOR, 2 AGAINST

Right wing male deputy: 4 FOR, 2 AGAINST

Right wing female deputy: 4 FOR, 2 AGAINST

Right wing male deputy: 4 FOR, 2 AGAINST

Left wing

Right wing

Diagram illustrating voting patterns for the Left wing. It shows two rows of two deputies each. Each deputy is represented by a name bar, a portrait, and a small colored square (green for female, red for male). To the right of each deputy are four document icons and a sign. The signs are either green with 'FOR' or red with 'AGAINST'. In the top row, the female deputy has 3 'FOR' and 1 'AGAINST' sign, while the male deputy has 1 'FOR' and 3 'AGAINST' signs. In the bottom row, the male deputy has 2 'FOR' and 2 'AGAINST' signs, while the female deputy has 2 'FOR' and 2 'AGAINST' signs.

Diagram illustrating voting patterns for the Right wing. It shows two rows of two deputies each. Each deputy is represented by a name bar, a portrait, and a small colored square (blue for female, black for male). To the right of each deputy are four document icons and a sign. The signs are either blue with 'AGAINST' or black with 'FOR'. In the top row, the female deputy has 3 'AGAINST' and 1 'FOR' sign, while the male deputy has 1 'AGAINST' and 3 'FOR' signs. In the bottom row, the male deputy has 2 'AGAINST' and 2 'FOR' signs, while the female deputy has 2 'AGAINST' and 2 'FOR' signs.

Pairwise agreement:  
25%

other  
deputies  
⋮

Left wing

Right wing

Fisheries ballots  
Pairwise agreement:  
100%

other  
deputies  
⋮



We introduce the problem of discovering **particular contexts** and **collections of individuals** where their **pairwise behavior** exceptionally differs from their usual one





We introduce the problem of discovering **particular contexts** and **collections of individuals** where their **pairwise behavior** exceptionally differs from their usual one



Find the **top-k three-set patterns**  $(c, g', g'')$  w.r.t. some quality measure  $\varphi$



We introduce the problem of discovering **particular contexts** and **collections of individuals** where their **pairwise behavior** exceptionally differs from their usual one



Find the **top-k three-set patterns**  $(c, g', g'')$  w.r.t. some quality measure  $\varphi$

Context

definition by **intent** of a **subset of items**



We introduce the problem of discovering **particular contexts** and **collections of individuals** where their **pairwise behavior** exceptionally differs from their usual one



Find the **top-k three-set patterns**  $(\underline{c}, g', g'')$  w.r.t. some quality measure  $\varphi$

Context

definition by **intent** of a **subset of items**

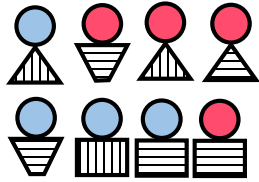
$g' \& g''$

definition by **intent** of a **subset of individuals**

- **DSC (Discovering similarities change) Framework**
- **DSC algorithm in a nutshell**
- **Examples**

- **DSC (Discovering similarities change) Framework**
- DSC algorithm in a nutshell
- Examples

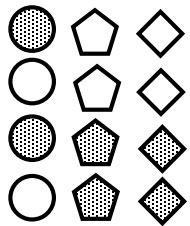
# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK



Reviewers  
*(eg. Users, Deputies)*

	○	◻	△	...
●	⌞	⌞		
●	⌞			
●	⌞	⌞	⌞	
⋮				⋮

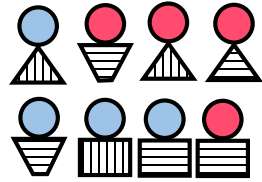
Reviews  
*(eg. Scores, Votes)*




Reviewees  
*(eg. Movies, Vote ballots)*

# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

Dataset example: Parliament voting dataset

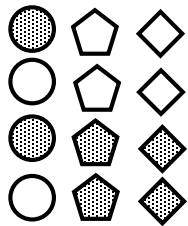


Reviewers  
(eg. Users, Deputies)



	○	■	△	...
●	⌞	⌞		
●	⌞			
●	⌞	⌞	⌞	
⋮				

Reviews  
(eg. Scores, Votes)

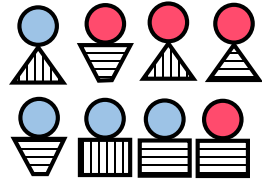


Reviewees  
(eg. Movies, Vote ballots)

<i>Items (Ballots) - E</i>			<i>Individuals (Deputies) - U</i>			<i>Outcome</i>
<i>Idsession</i>	<i>Date</i>	<i>Theme</i>	<i>Full name</i>	<i>National Party</i>	<i>Political Group</i>	<i>Vote</i>
001	2017/03/17	1.10 Justice 2.10 Europe coop	Lavrilleux	LR	PPE	For
001	2017/03/17	1.10 Justice 2.10 Europe coop	Philippot	FN	ENF	Against
002	2017/04/11	3.10 Agriculture	Lavrilleux	LR	PPE	For
002	2017/04/11	3.10 Agriculture	Philippot	FN	ENF	For
002	2017/04/11	3.10 Agriculture	Arnatu	FN	ENF	For
003	2017/04/11	1.20 Security	Le Grip	LR	PPE	Abstain

# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

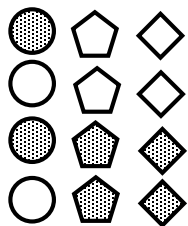
Dataset example: Parliament voting dataset



Reviewers  
(eg. Users, Deputies)

	○	■	△	...
●	⌞	⌞		
●	⌞			
●	⌞	⌞	⌞	

Reviews  
(eg. Scores, Votes)



Reviewees  
(eg. Movies, Vote ballots)

Items (Ballots) - E			Individuals (Deputies) - U			Outcome
Idsession	Date	Theme	Full name	National Party	Political Group	Vote
001	2017/03/17	1.10 Justice 2.10 Europe coop	Lavrilleux	LR	PPE	For
001	2017/03/17	1.10 Justice 2.10 Europe coop	Philippot	FN	ENF	Against
002	2017/04/11	3.10 Agriculture	Lavrilleux	LR	PPE	For
002	2017/04/11	3.10 Agriculture	Philippot	FN	ENF	For
002	2017/04/11	3.10 Agriculture	Arnatu	FN	ENF	For
003	2017/04/11	1.20 Security	Le Grip	LR	PPE	Abstain

Descriptions attributes\* over  
items (context)

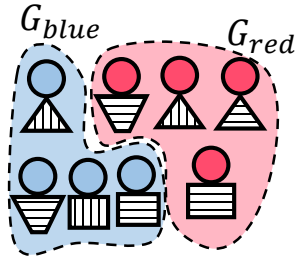
Descriptions attributes\* over  
individuals

\*numeric, nominal, hierarchical multi-tag attributes



# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

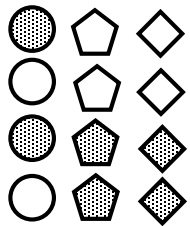
**1** *Constitute groups*  
(eg. By head color)



Reviewers  
(eg. Users, Deputies)

	○	■	△	...
●	⌞	⌞		
●	⌞			
●	⌞	⌞	⌞	
⋮				⋮

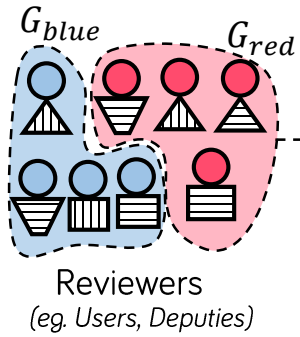
Reviews  
(eg. Scores, Votes)



Reviewees  
(eg. Movies, Vote ballots)

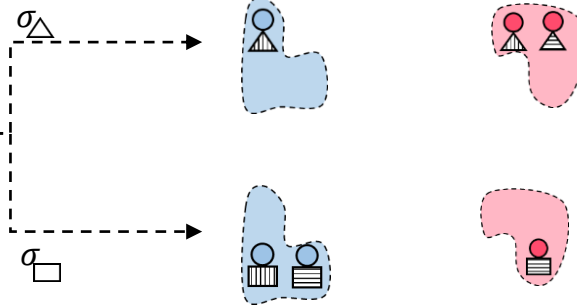
# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

**1** *Constitute groups*  
(eg. By head color)



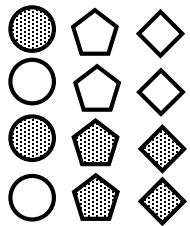
**2**

Generate a particular couple of subsets of users  
eg. Confront  $\square$  vs.  $\triangle$



	○	◻	△	...
●	◻	◻		
●	◻			
●	◻	◻	◻	
⋮				⋮

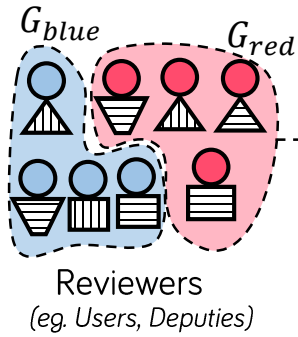
Reviews  
(eg. Scores, Votes)



Reviewees  
(eg. Movies, Vote ballots)

# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

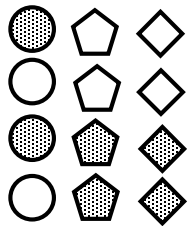
**1** *Constitute groups*  
(eg. By head color)



**2** *Generate a particular couple of subsets of users*  
eg. Confront  $\square$  vs.  $\triangle$

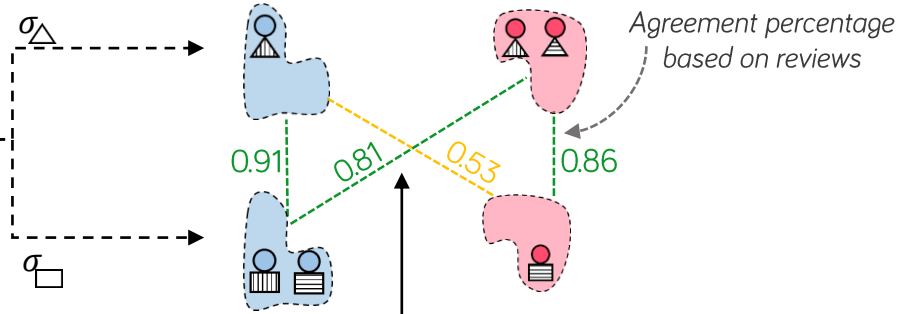
	○	◻	△	...
●	◻	△		
○	◻			
●	◻	△		
○	◻	△		

Reviews  
(eg. Scores, Votes)



Reviewees  
(eg. Movies, Vote ballots)

**3** *Global pairwise behavior*

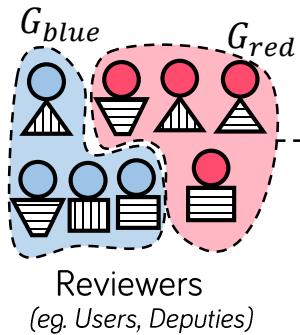


Consider all reviewees

# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

D  
S  
C  
  
O  
V  
E  
R  
V  
I  
E  
W

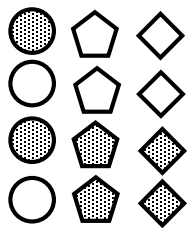
**1** *Constitute groups*  
(eg. By head color)



**2** *Generate a particular couple of subsets of users*  
eg. Confront  $\square$  vs.  $\triangle$

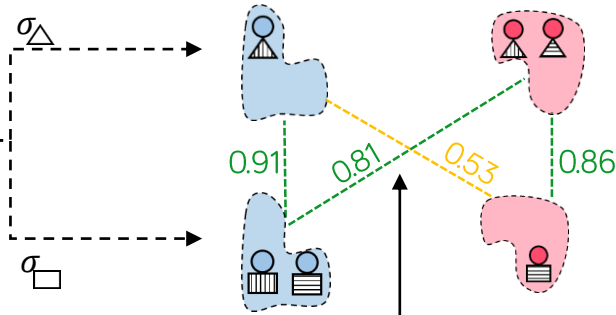
	○	◻	△	...
●	◻	△		
○	◻			
●	◻	△		
○	◻	△		

Reviews  
(eg. Scores, Votes)



Reviewees  
(eg. Movies, Vote ballots)

**3** *Global pairwise behavior*

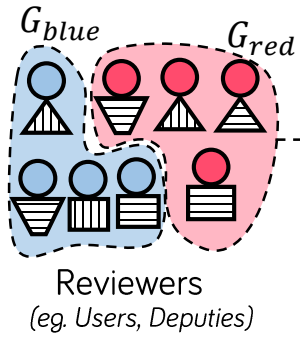


	●	●
●	0.91	0.81
●	0.53	0.86

Consider all reviewees

# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

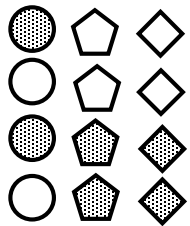
**1** *Constitute groups*  
(eg. By head color)



**2** *Generate a particular couple of subsets of users*  
eg. Confront  $\square$  vs.  $\triangle$

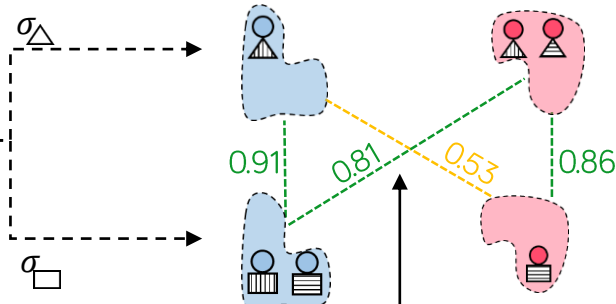
	○	◻	△	...
●	◻	△		
◻	△			
△				
...				

Reviews  
(eg. Scores, Votes)

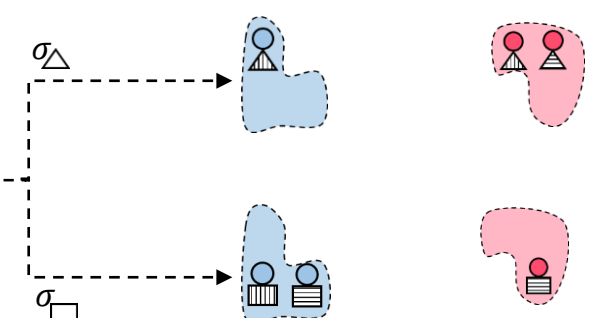


Reviewees  
(eg. Movies, Vote ballots)

**3** *Global pairwise behavior*



Consider all reviewees

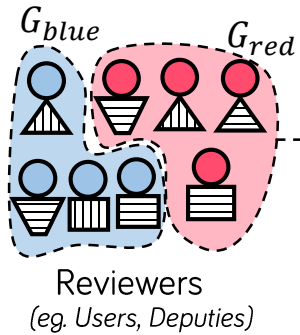


	0.91	0.81
	0.53	0.86

# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

D  
S  
C  
  
O  
V  
E  
R  
V  
I  
E  
W

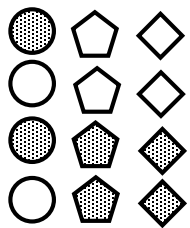
**1** *Constitute groups*  
(eg. By head color)



**2** *Generate a particular couple of subsets of users*  
eg. Confront  $\square$  vs.  $\triangle$

	○	◻	△	...
●	◻	△		
●	◻			
●	◻	△		
...				

Reviews  
(eg. Scores, Votes)

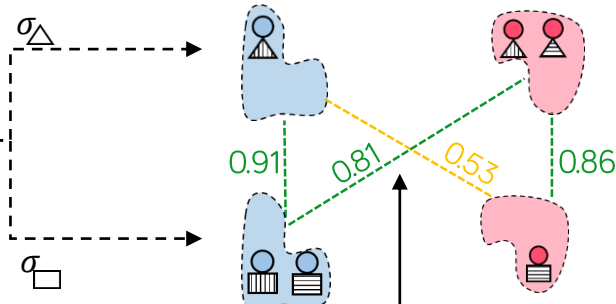


Reviewees  
(eg. Movies, Vote ballots)

Consider all reviewees

**4** *Generate a subset of reviewees*  
eg. Dotted diamonds

**3** *Global pairwise behavior*

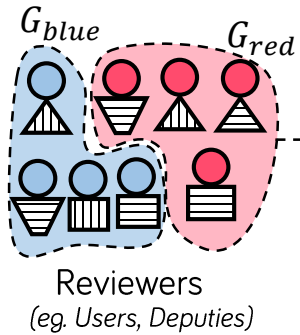


	0.91	0.81
	0.53	0.86

# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

DSC OVERVIEW

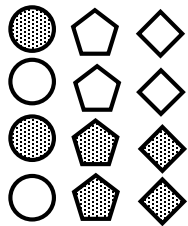
**1** Constitute groups  
(eg. By head color)



**2** Generate a particular couple of subsets of users  
eg. Confront  $\square$  vs.  $\triangle$

	$\circ$	$\square$	$\triangle$	...
$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	
$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	
$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	

Reviews  
(eg. Scores, Votes)

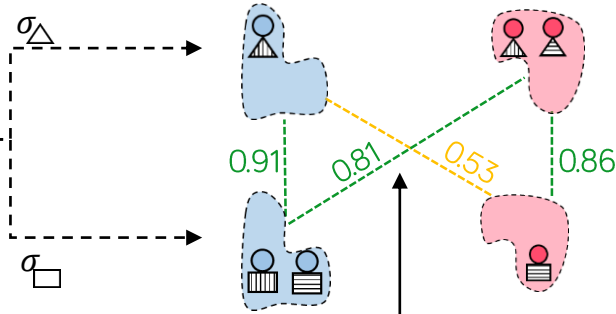


Reviewees  
(eg. Movies, Vote ballots)

**4** Generate a subset of reviewees  
eg. Dotted diamonds  $\blacklozenge$

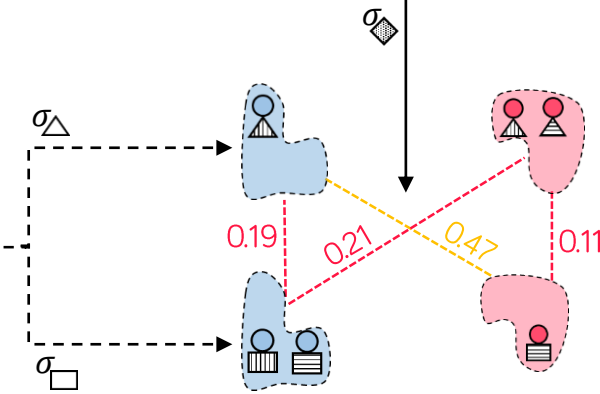
Consider all reviewees

**3** Global pairwise behavior



$\text{reviewer icon}$	$\text{reviewer icon}$
0.91	0.81
0.53	0.86

**5** Contextual pairwise behavior

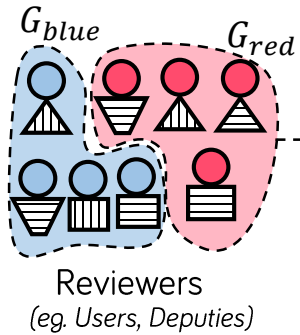


$\text{reviewer icon}$	$\text{reviewer icon}$
0.19	0.21
0.47	0.11

# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

DSC OVERVIEW

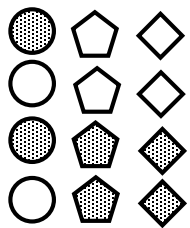
**1** *Constitute groups*  
(eg. By head color)



**2** *Generate a particular couple of subsets of users*  
eg. Confront  $\square$  vs.  $\triangle$

	$\circ$	$\square$	$\triangle$	...
$\bullet$	$\uparrow$	$\uparrow$		
$\square$	$\uparrow$			
$\triangle$	$\uparrow$	$\uparrow$		
...				

*Reviews*  
(eg. Scores, Votes)

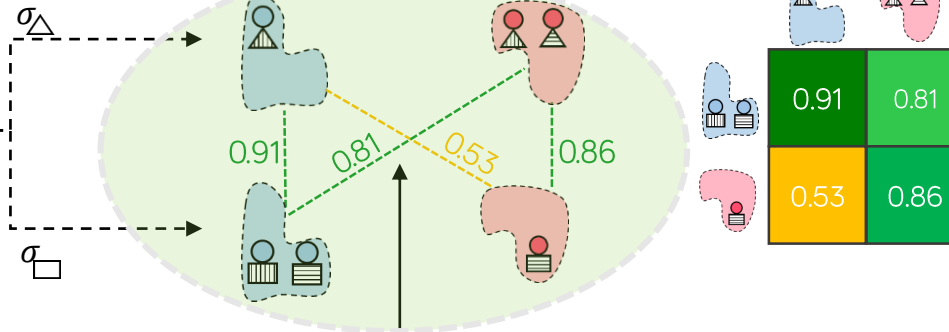


*Reviewees*  
(eg. Movies, Vote ballots)

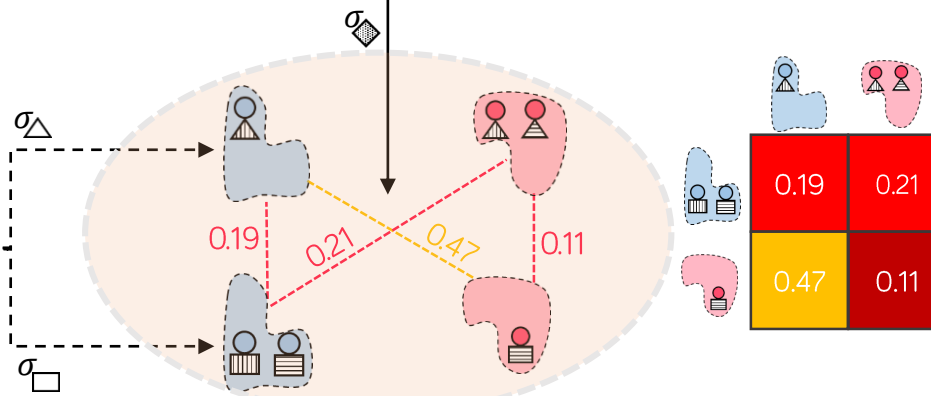
**4** *Generate a subset of reviewees*  
eg. Dotted diamonds  $\blacklozenge$

Consider all reviewees

**3** *Global pairwise behavior*



**5** *Contextual pairwise behavior*



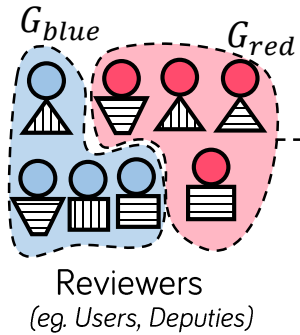


# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK

DSC OVERVIEW

4

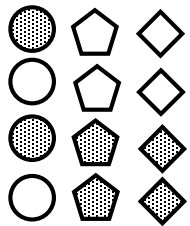
**1** Constitute groups  
(eg. By head color)



**2** Generate a particular couple of subsets of users  
eg. Confront  $\square$  vs.  $\triangle$

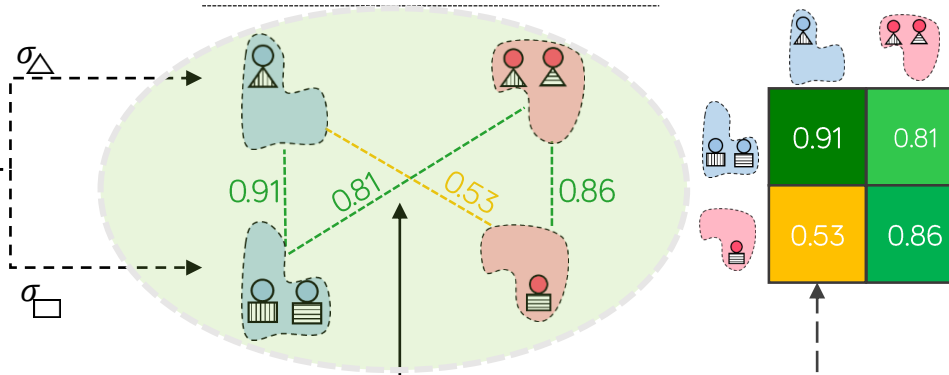
	$\circ$	$\square$	$\triangle$	...
$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	
$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	
$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	$\text{reviewer icon}$	

Reviews  
(eg. Scores, Votes)



**4** Generate a subset of reviewees  
eg. Dotted diamonds  $\blacklozenge$

**3** Global pairwise behavior

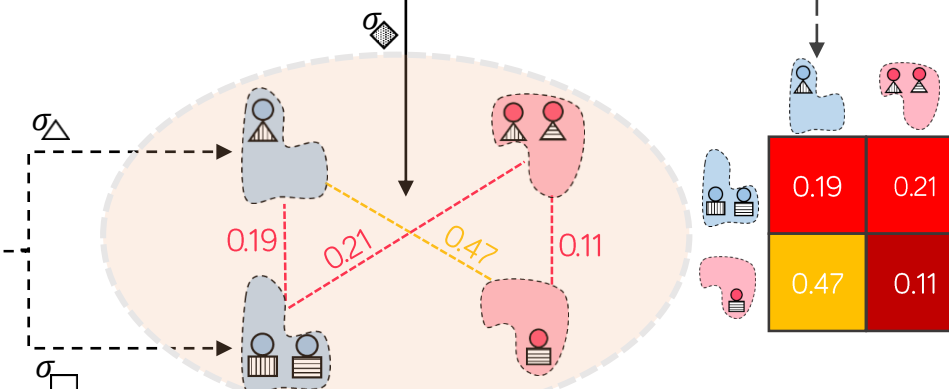


Consider all reviewees

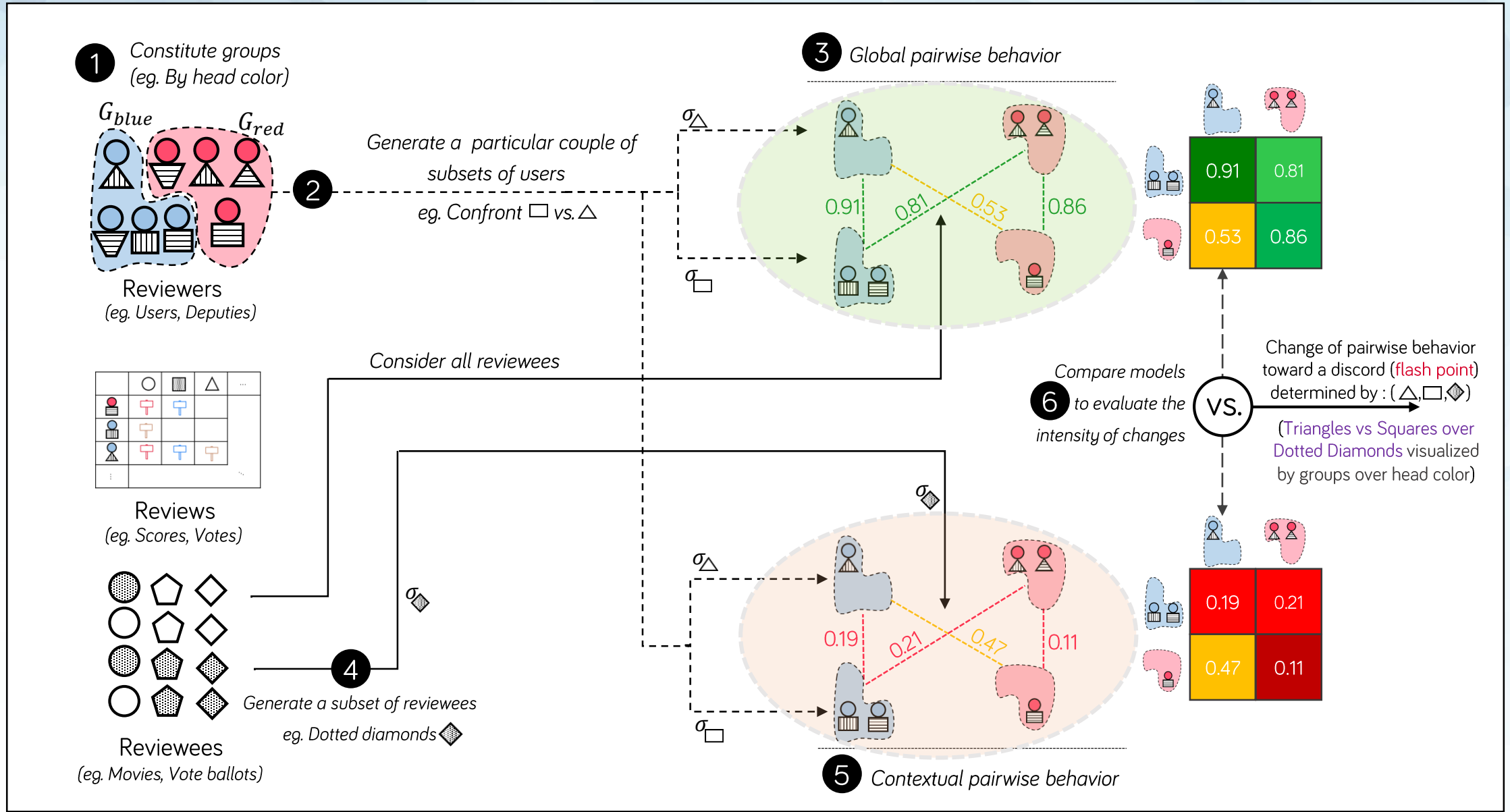
**6** Compare models  
to evaluate the intensity of changes

VS.

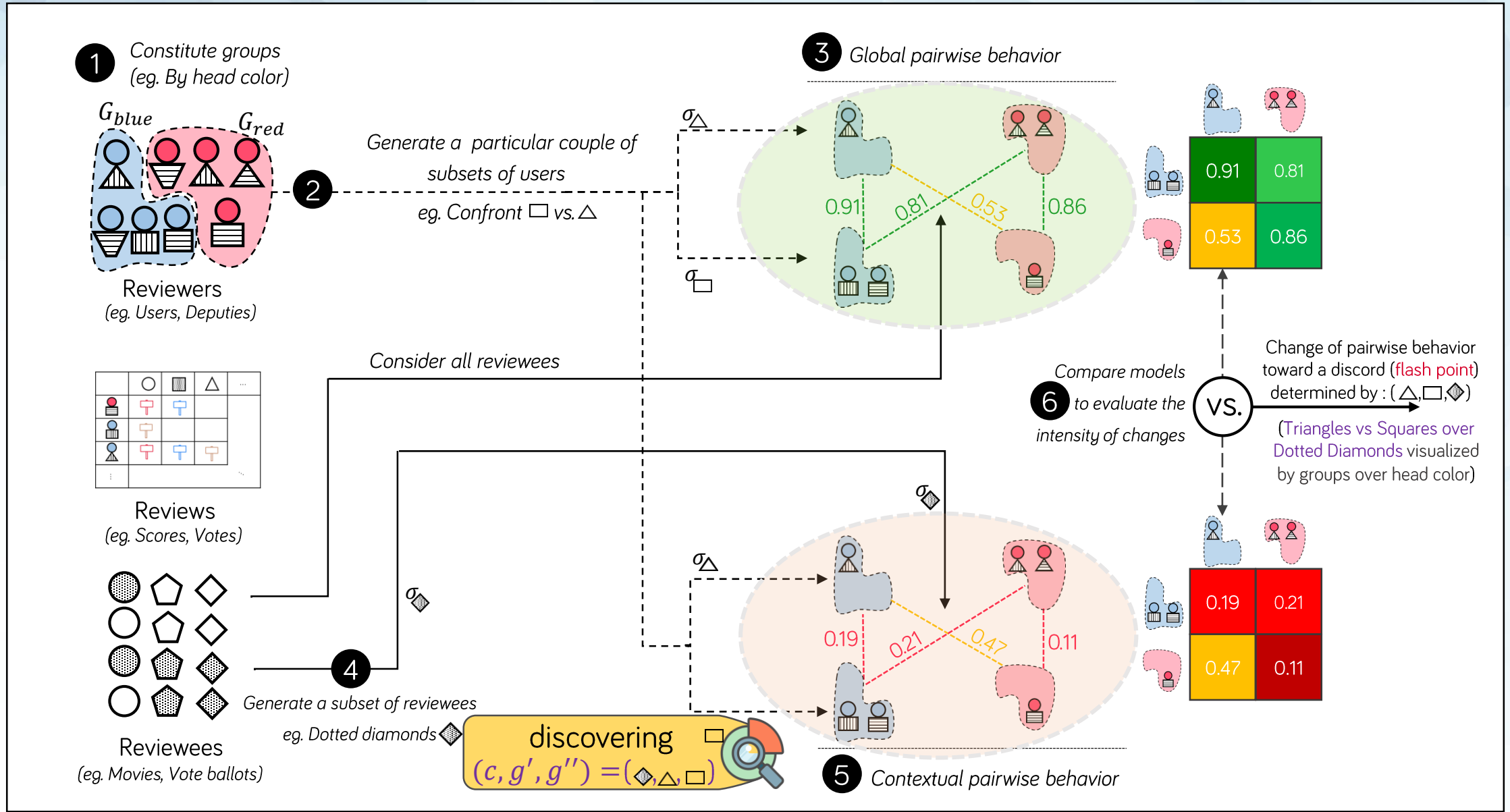
**5** Contextual pairwise behavior



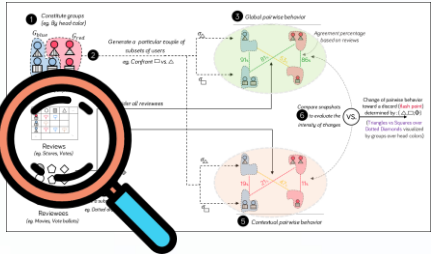
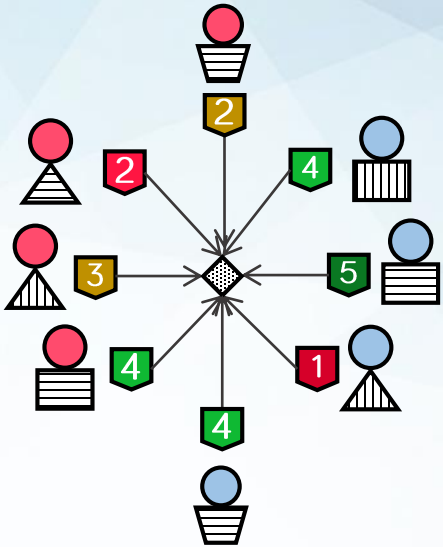
# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK



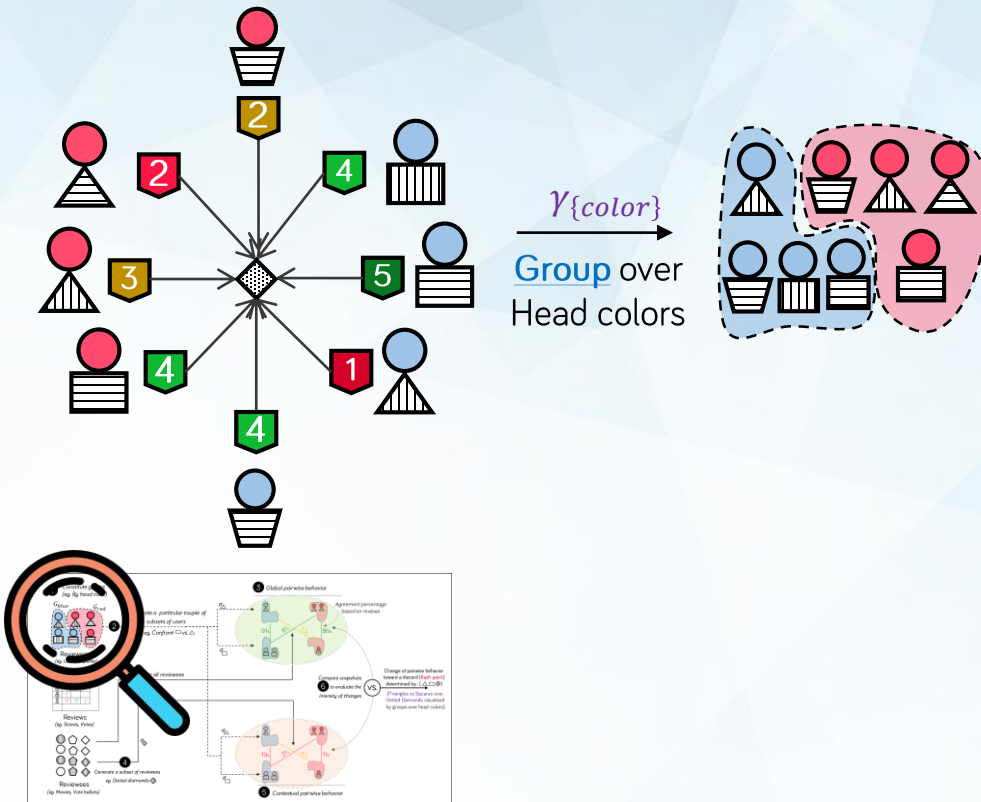
# DISCOVERING SIMILARITIES CHANGE ( DSC ) FRAMEWORK



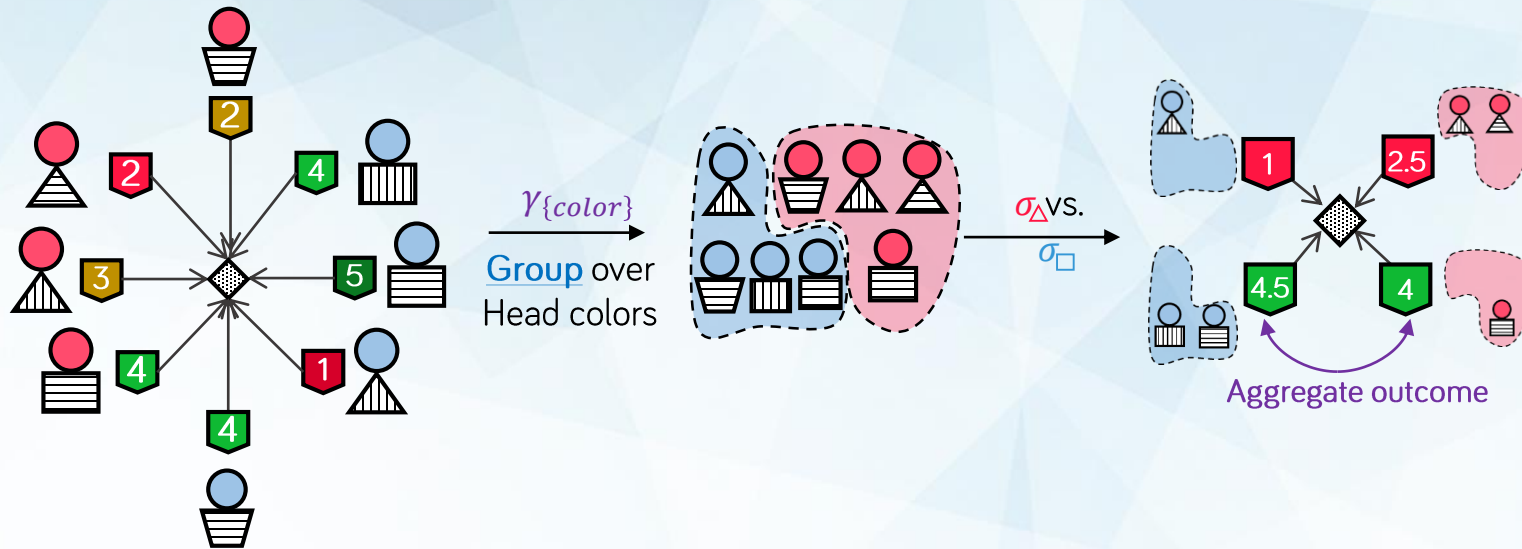
Individuals described by (color, shapes)



Individuals described by (color, shapes)



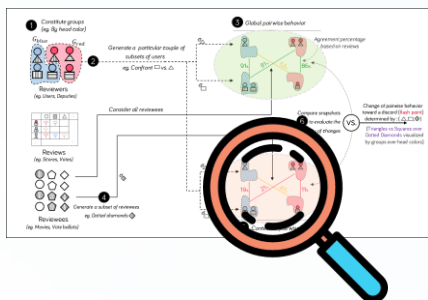
Individuals described by (color, shapes)



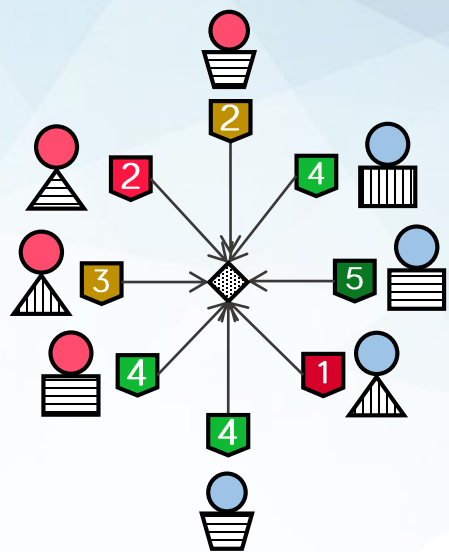
Several aggregation operators, e.g.:

Majorities votes:  $\theta(e, G) = \operatorname{argmax}_{v \in O} \operatorname{count}(v, \{\operatorname{outcome}(e, u) | u \in G\})$

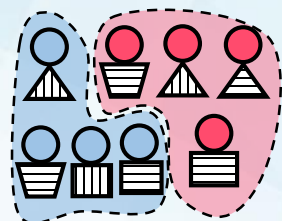
Rating average:  $\theta(e, G) = \frac{1}{|G|} \sum_{u \in G} \operatorname{outcome}(e, u)$



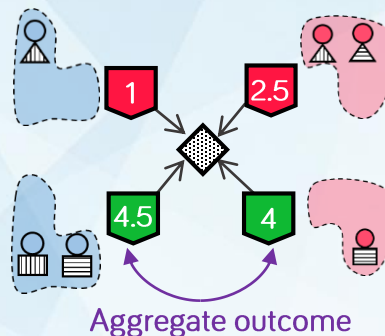
Individuals described by (color, shapes)



$\gamma_{\{color\}}$   
Group over  
Head colors

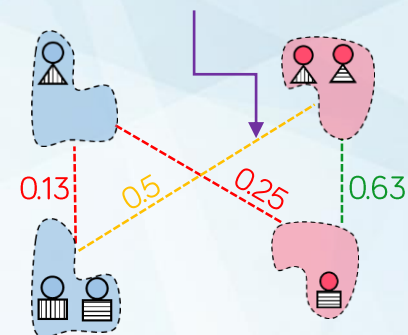


$\sigma_{\Delta}$  vs.  
 $\sigma_{\square}$



Compute similarities  
between groups

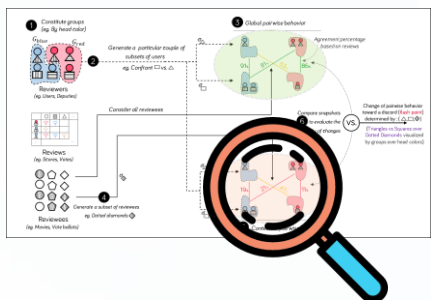
$Simobj(\diamond, red\Delta, blue\square)$



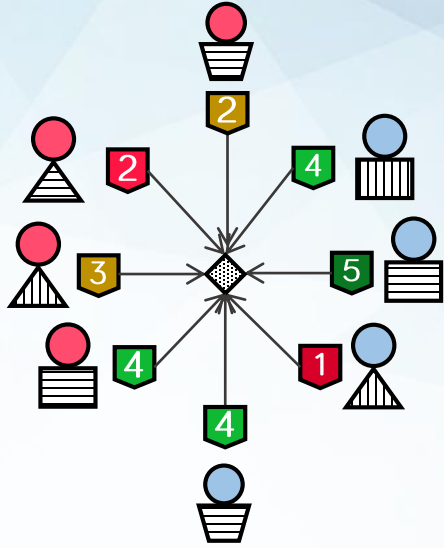
Similarity functions over one item:  $simobj$

Rating agreement:  $simobj(e, G_1, G_2) = 1 - \frac{1}{scaleSize} |\theta(e, G_1) - \theta(e, G_2)|$

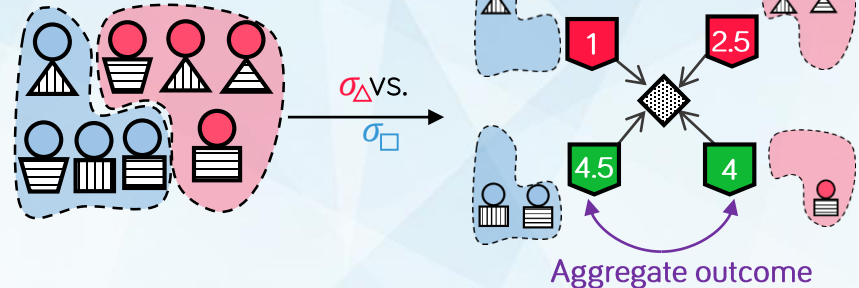
Voting similarity:  $simobj(e, G_1, G_2) = \delta_{\theta(e, G_1)\theta(e, G_2)}$



Individuals described by (color, shapes)

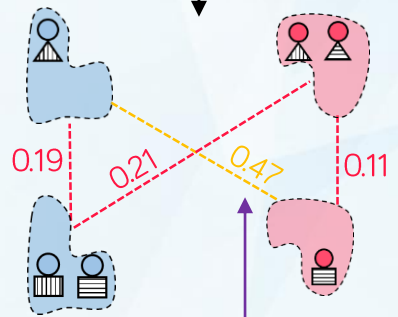
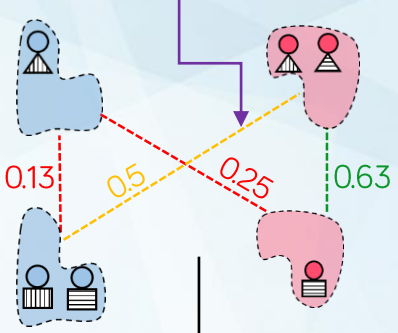


$\gamma_{\{color\}}$   
Group over  
Head colors



Compute similarities  
between groups

$Simobj(\blacklozenge, red\Delta, blue\square)$

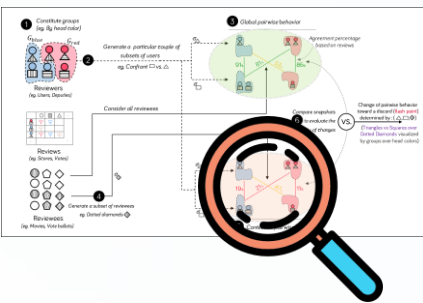


$sim(\blacktriangle, red\Delta, blue\square)$

some items

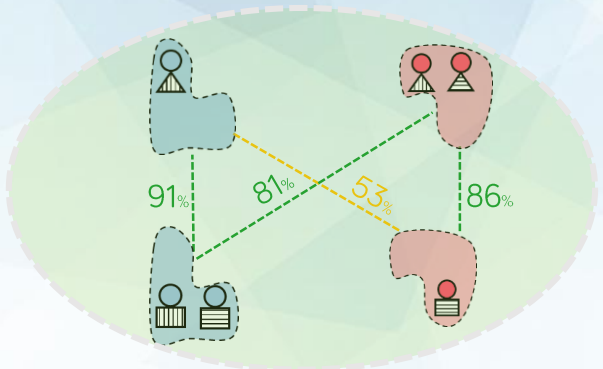


Similarity over a set of items: *sim* is depicted by an average (or a weighted average) of *simobj* over a given set of items

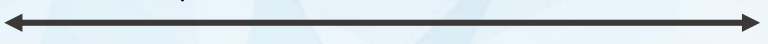




Global pairwise behavior

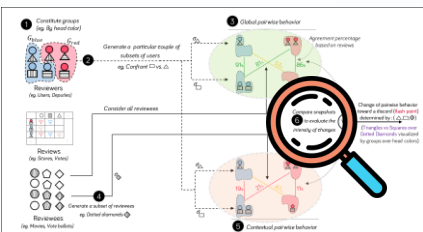
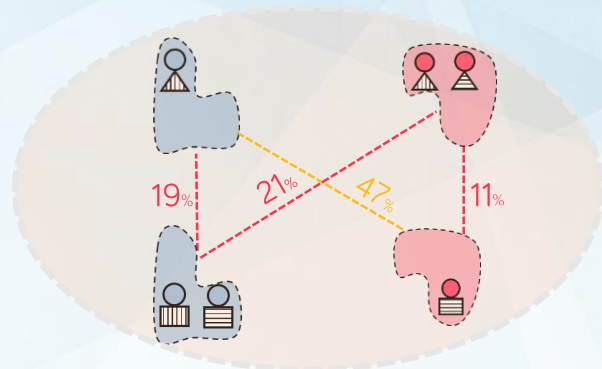


Comparison between models

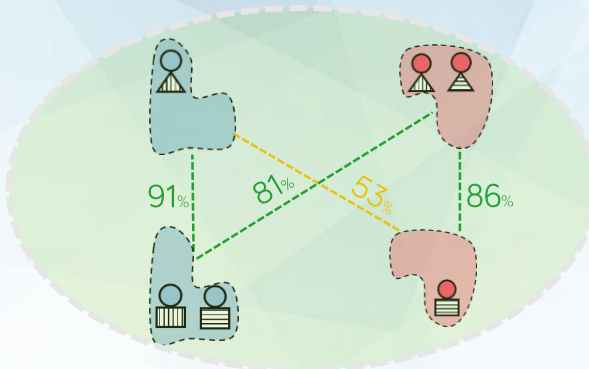


Use of quality measure

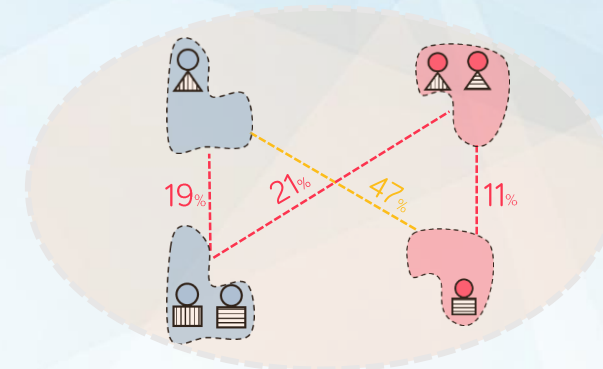
Contextual pairwise behavior



Global pairwise behavior



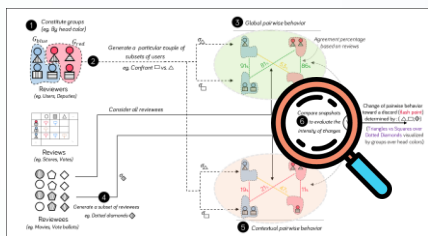
Contextual pairwise behavior



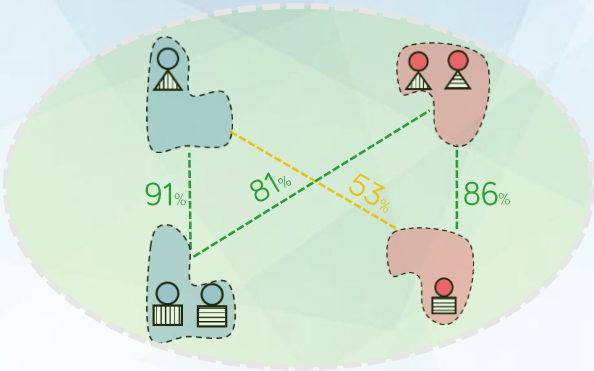
Comparison between models  
Use of quality measure

$\Phi_{consent}$ : to uncover contexts that leads to more agreement

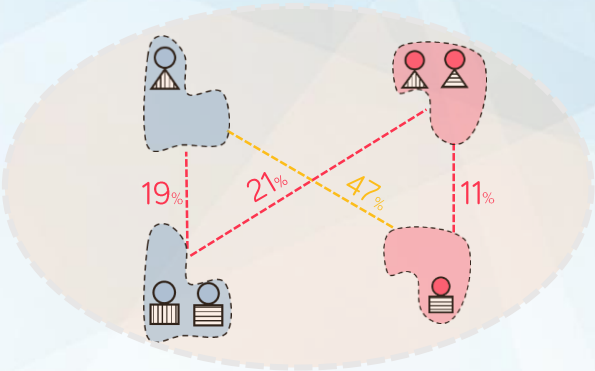
$$\Phi_{consent} = \frac{\sum_{(i,j) \in \gamma_L(U_{gI}) \times \gamma_L(U_{gII})} \max(\text{sim}(E_{context,i,j}) - \text{sim}(E,i,j), 0)}{|\gamma_L(U_{gI})| \times |\gamma_L(U_{gII})|}$$



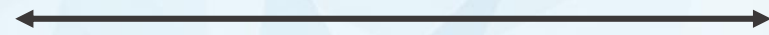
Global pairwise behavior



Contextual pairwise behavior



Comparison between models



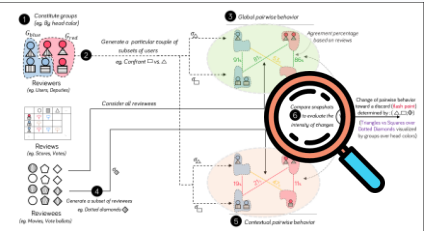
Use of quality measure



$\varphi_{consent}$ : to uncover contexts that leads to more agreement

$\varphi_{dissent}$ : to uncover contexts that leads to more disagreement

$$\varphi_{dissent} = \frac{\sum_{(i,j) \in \gamma_L(U_{gI}) \times \gamma_L(U_{gII})} \max(\text{sim}(E, i, j) - \text{sim}(E_{context}, i, j), 0)}{|\gamma_L(U_{gI})| \times |\gamma_L(U_{gII})|}$$



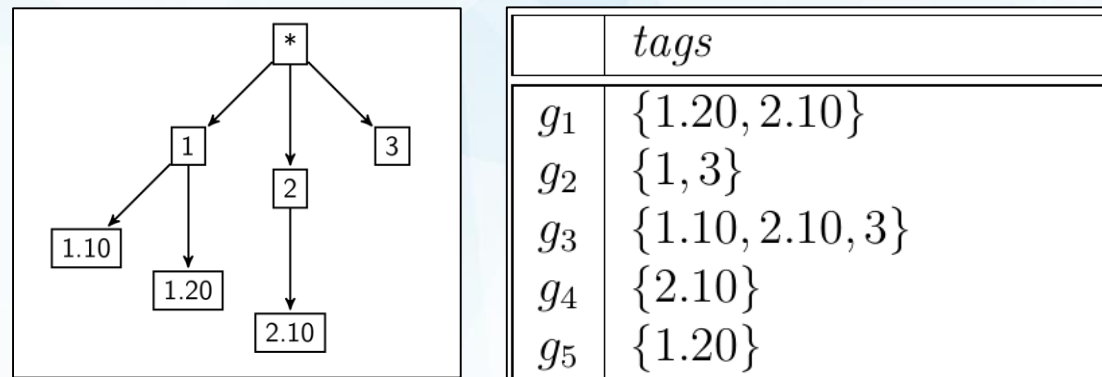
- DSC (Discovering similarities change) Framework
- **DSC algorithm in a nutshell**
- Examples

a) **Branch and bound algorithm:** definition of upper bounds on both quality measures for a more efficient enumeration.

a) **Branch and bound algorithm:** definition of upper bounds on both quality measures for a more efficient enumeration.

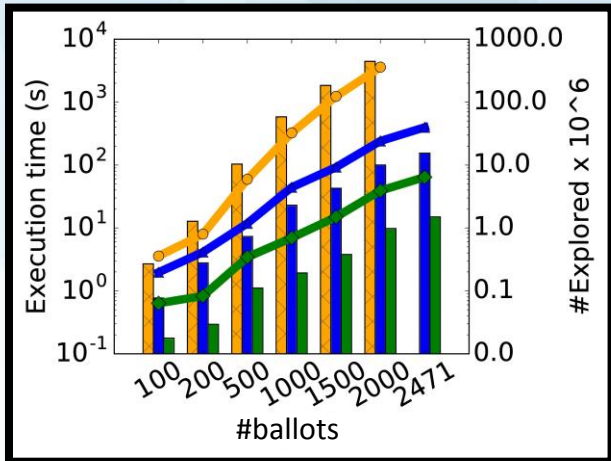
b) **Closed descriptions:** reduce considerably the search space by relying on closed pattern mining.

\* **New pattern domain:** HMT (**Hierarchical Multi-tag attribute**) relying on pattern structures<sup>†</sup>



<sup>†</sup> Ganter, B., & Kuznetsov, S. (2001). Pattern structures and their projections. *Conceptual Structures: Broadening the Base*, 129-142

# DSC PERFORMANCE



European parliament datasets



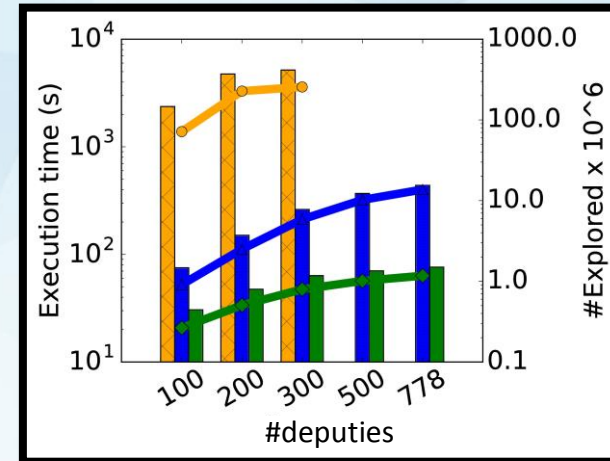
784 Deputies



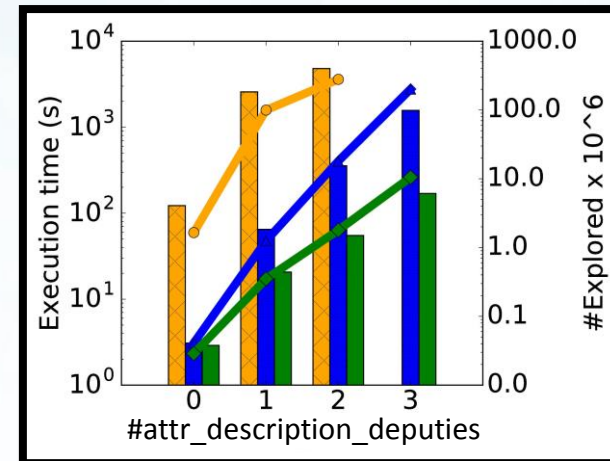
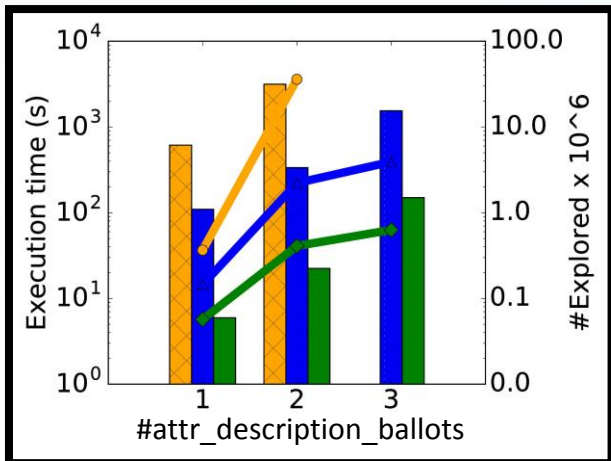
2471 ballots



1 600 000 votes

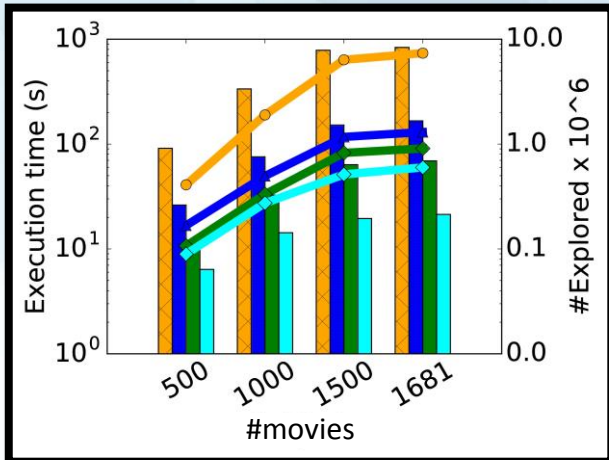


- Baseline
- Closed
- Branch & bound + Closed DSC (UB1)



- Execution time
- #Explored descriptions

# DSC PERFORMANCE



Movielens dataset



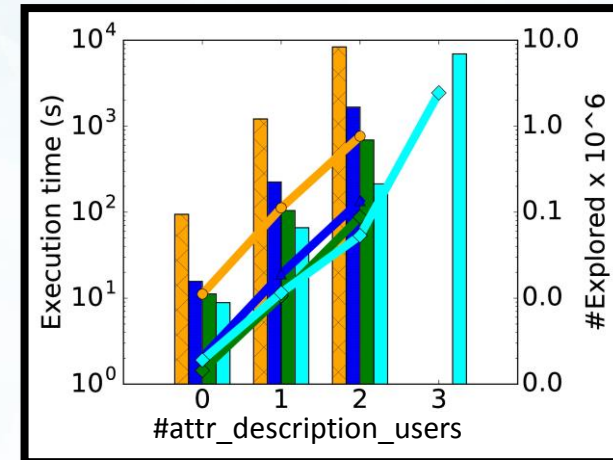
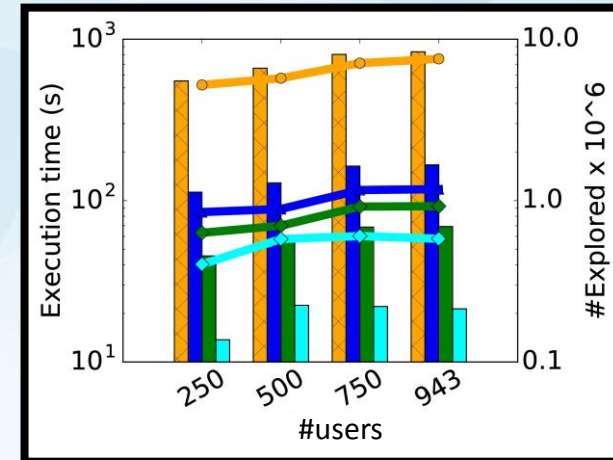
943 Users



1681 Movies



100 000 ratings

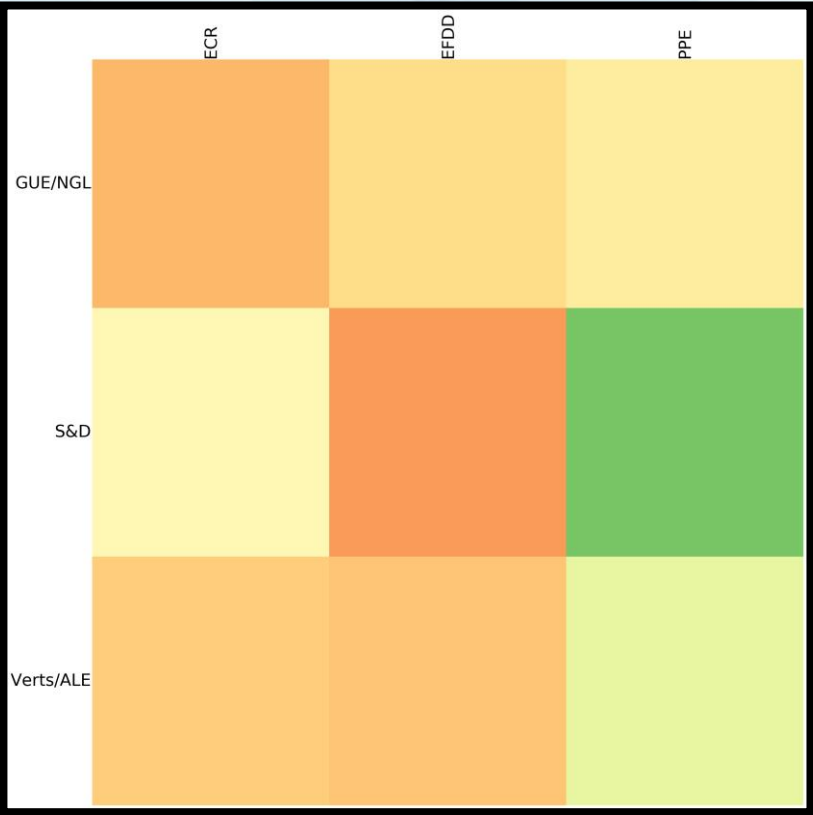


- Baseline
- Closed
- Branch & bound + Closed DSC (UB1)
- Branch & bound + Closed DSC (UB2)
- Execution time
- #Explored descriptions

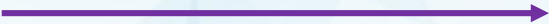


- DSC (Discovering similarities change) Framework
- Algorithm in a nutshell
- **Examples**

## Usual pairwise behavior

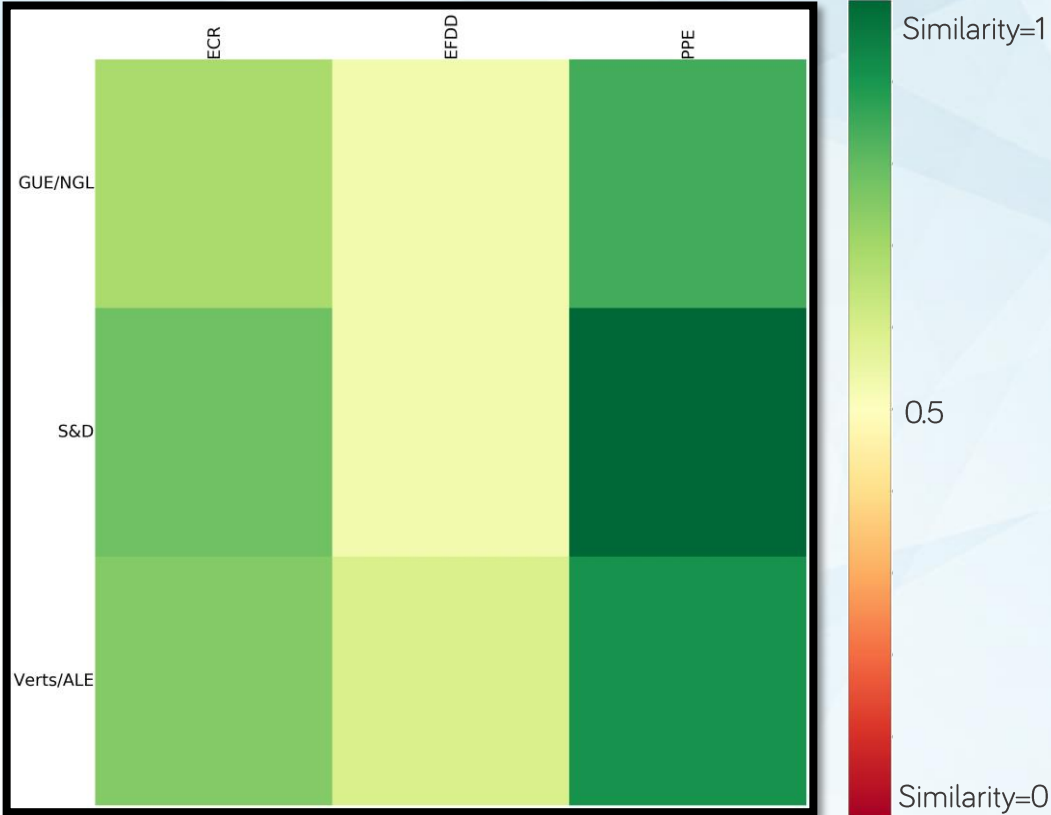


Toward **consent** between European political groups  
The pattern:

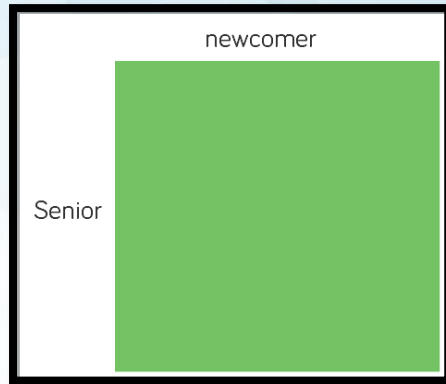


[7.40 European judicial conventions during Feb. – Nov. 2015  
,left wing ,right wing]

## Contextual pairwise behavior



## Usual pairwise behavior

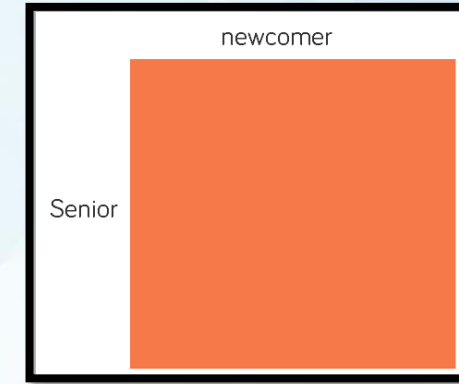


Toward **Dissent** between Yelp Users for the context:



[Professional Services, Shopping, In Oklahoma, Senior, Newcomer]

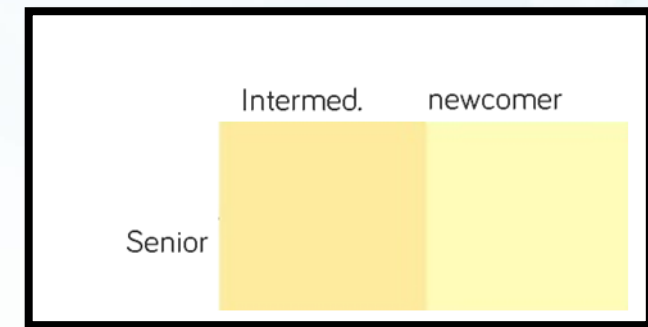
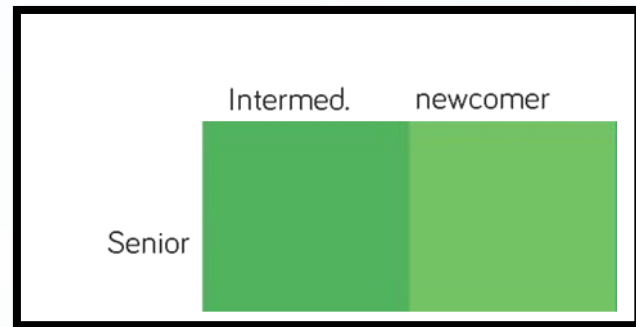
## Contextual pairwise behavior



Toward **Dissent** between Yelp Users for the context:



[Medical Center, Doctors, In Wisconsin, {Senior}, {Intermed., Newcomer},]



## Conclusion



- ◇ Definition of a novel problem :  
discovering exceptional pairwise behavior
- ◇ Implementation of a **branch and bound**  
enumeration algorithm (DSC)
- ◇ Experimentation over **real-world**  
**datasets** (voting and rating)

## Conclusion



- ◆ Definition of a novel problem : discovering exceptional pairwise behavior
- ◆ Implementation of a branch and bound enumeration algorithm (DSC)
- ◆ Experimentation over real-world datasets (voting and rating)



## Perspectives

- ◆ ContentCheck: Offering a set of tools for journalists (Le Monde) of lead finding/fact checking
- ◆ Providing adapted instant mining and interactive mining algorithms.

THANKS

FOR YOUR TIME



QUESTIONS

Contact : [adnene.belfodil@insa-lyon.fr](mailto:adnene.belfodil@insa-lyon.fr)

Materials: <https://github.com/Adnene93/DiscoveringSimilarityChanges>

*Looking forward to meet up with you during the poster session*



# BIBLIOGRAPHY

THANKS  
FOR YOUR TIME



- ⑥ W. Duivesteijn, A. J. Feelders, and A. Knobbe. Exceptional model mining. *Data Mining and Knowledge Discovery*, 30(1):47{98, 2016.
- ⑥ B. Ganter and S. O. Kuznetsov. Pattern structures and their projections. ICCS,2001.
- ⑥ C. R. de Sà, W. Duivesteijn, C. Soares, and A. Knobbe. Exceptional preferences mining. In *DS*, pages 3{18. Springer, 2016.
- ⑥ G. Bosc, J. Golebiowski, M. Bensa, C. Robardet, M. Plantevit, J. Boulicaut, and M. Kaytoue. Local subgroup discovery for eliciting and understanding new structure-odor relationships. In *DS*, pages 19{34, 2016.
- ⑥ S. Amer-Yahia, S. Kleisarchaki, N. K. Kolloju, L. V. Lakshmanan, and R. H. Zamar. Exploring rated datasets with rating maps. In *WWW*, 2017.
- ⑥ M. Das, S. Amer-Yahia, G. Das, and C. Yu. Mri: Meaningful interpretations of collaborative ratings. *PVLDB*

**Contact :** [adnene.belfodil@insa-lyon.fr](mailto:adnene.belfodil@insa-lyon.fr)

**Materials:** <https://github.com/Adnene93/DiscoveringSimilarityChanges>

*Looking forward to meet up with you during the poster session*