



DEvIANT: DISCOVERING SIGNIFICANT EXCEPTIONAL (DIS-)AGREEMENT WITHIN GROUPS

AUTHORS. ● Adnene Belfodil ● Wouter Duivesteijn

● Marc Plantevit ● Sylvie Cazalens ● Philippe Lamarre



INTRODUCTION

PROBLEM DEFINITION

Data (Behavioral Data)



Individuals (Parliamentarians)

<i>idi</i>	Party	State	Age	
<i>i</i> ₁	Democrats		New York	32
<i>i</i> ₂	Democrats		Florida	41
<i>i</i> ₃	Republicans		Texas	45
<i>i</i> ₄	Republicans		Chicago	39
<i>i</i> ₅	Republicans		Washington	28
<i>i</i> ₆	Republicans		Illinois	51



U.S. House of Representatives voting sessions

Data (Behavioral Data)



Individuals (Parliamentarians)

<i>idi</i>	<i>Party</i>	<i>State</i>	<i>Age</i>
<i>i</i> ₁	Democrats	New York	32
<i>i</i> ₂	Democrats	Florida	41
<i>i</i> ₃	Republicans	Texas	45
<i>i</i> ₄	Republicans	Chicago	39
<i>i</i> ₅	Republicans	Washington	28
<i>i</i> ₆	Republicans	Illinois	51



Entities (voting sessions)

<i>ide</i>	<i>Date</i>	<i>Themes</i>
<i>e</i> ₁	2016/04/16	1.05 National Budget
<i>e</i> ₂	2016/05/16	5.03 Employee benefits
<i>e</i> ₃	2016/06/16	7 Environment
<i>e</i> ₄	2016/06/16	1.20 Citizen's rights; 20.11 Gov. Adm. Issues
<i>e</i> ₅	2016/07/03	20.11 Gov. Adm. Issues
<i>e</i> ₆	2016/07/29	20.11 Gov. Adm. Issues



U.S. House of Representatives voting sessions

Data (Behavioral Data)



Individuals (Parliamentarians)			
<i>idi</i>	Party	State	Age
<i>i</i> ₁	Democrats		New York 32
<i>i</i> ₂	Democrats		Florida 41
<i>i</i> ₃	Republicans		Texas 45
<i>i</i> ₄	Republicans		Chicago 39
<i>i</i> ₅	Republicans		Washington 28
<i>i</i> ₆	Republicans		Illinois 51

<i>BD</i>	<i>e</i> ₁	<i>e</i> ₂	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆
<i>i</i> ₁						
<i>i</i> ₂						
<i>i</i> ₃						
<i>i</i> ₄						
<i>i</i> ₅						
<i>i</i> ₆						

Entities (voting sessions)		
<i>ide</i>	Date	Themes
<i>e</i> ₁	2016/04/16	1.05 National Budget
<i>e</i> ₂	2016/05/16	5.03 Employee benefits
<i>e</i> ₃	2016/06/16	7 Environment
<i>e</i> ₄	2016/06/16	1.20 Citizen's rights; 20.11 Gov. Adm. Issues
<i>e</i> ₅	2016/07/03	20.11 Gov. Adm. Issues
<i>e</i> ₆	2016/07/29	20.11 Gov. Adm. Issues



U.S. House of Representatives voting sessions

Groups

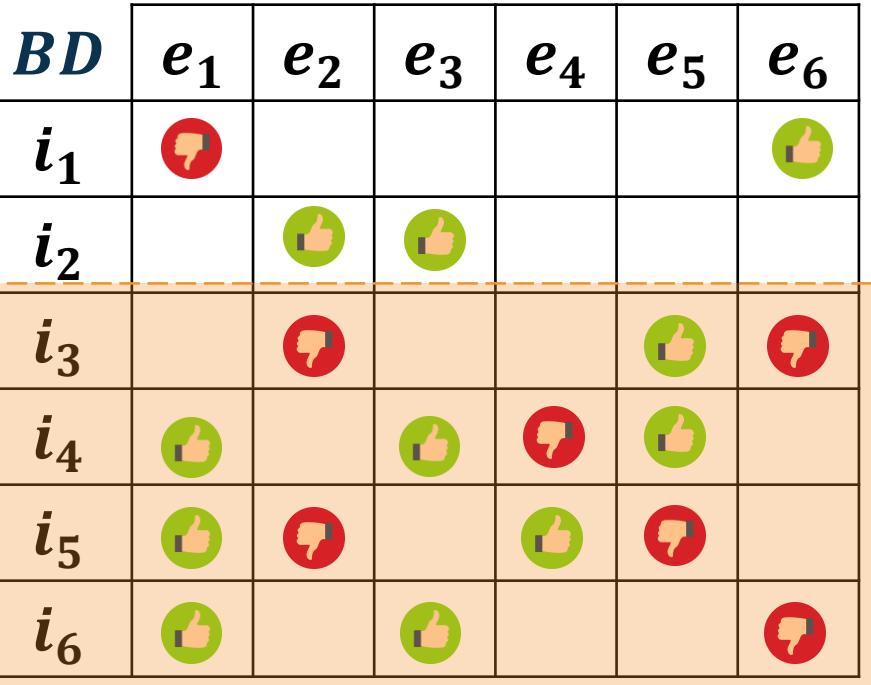


Individuals (Parliamentarians)

<i>idi</i>	Party	State	Age
<i>i</i> ₁	Democrats	New York	32
<i>i</i> ₂	Democrats	Florida	41
<i>i</i> ₃	Republicans	Texas	45
<i>i</i> ₄	Republicans	Chicago	39
<i>i</i> ₅	Republicans	Washington	28
<i>i</i> ₆	Republicans	Illinois	51

Party = Republicans

Group



Entities (voting sessions)

<i>ide</i>	Date	Themes
<i>e</i> ₁	2016/04/16	1.05 National Budget
<i>e</i> ₂	2016/05/16	5.03 Employee benefits
<i>e</i> ₃	2016/06/16	7 Environment
<i>e</i> ₄	2016/06/16	1.20 Citizen's rights; 20.11 Gov. Adm. Issues
<i>e</i> ₅	2016/07/03	20.11 Gov. Adm. Issues
<i>e</i> ₆	2016/07/29	20.11 Gov. Adm. Issues

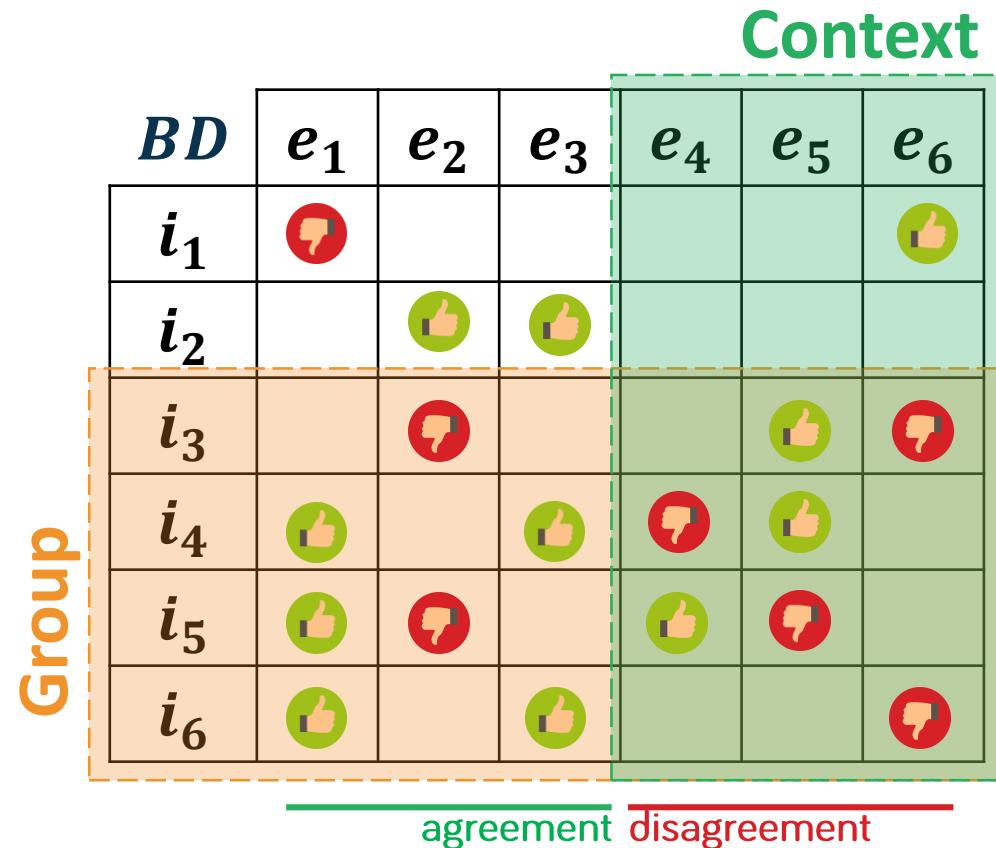
Groups and Contexts



Individuals (Parliamentarians)

<i>idi</i>	Party	State	Age	
<i>i</i> ₁	Democrats		New York	32
<i>i</i> ₂	Democrats		Florida	41
<i>i</i> ₃	Republicans		Texas	45
<i>i</i> ₄	Republicans		Chicago	39
<i>i</i> ₅	Republicans		Washington	28
<i>i</i> ₆	Republicans		Illinois	51

Party = Republicans

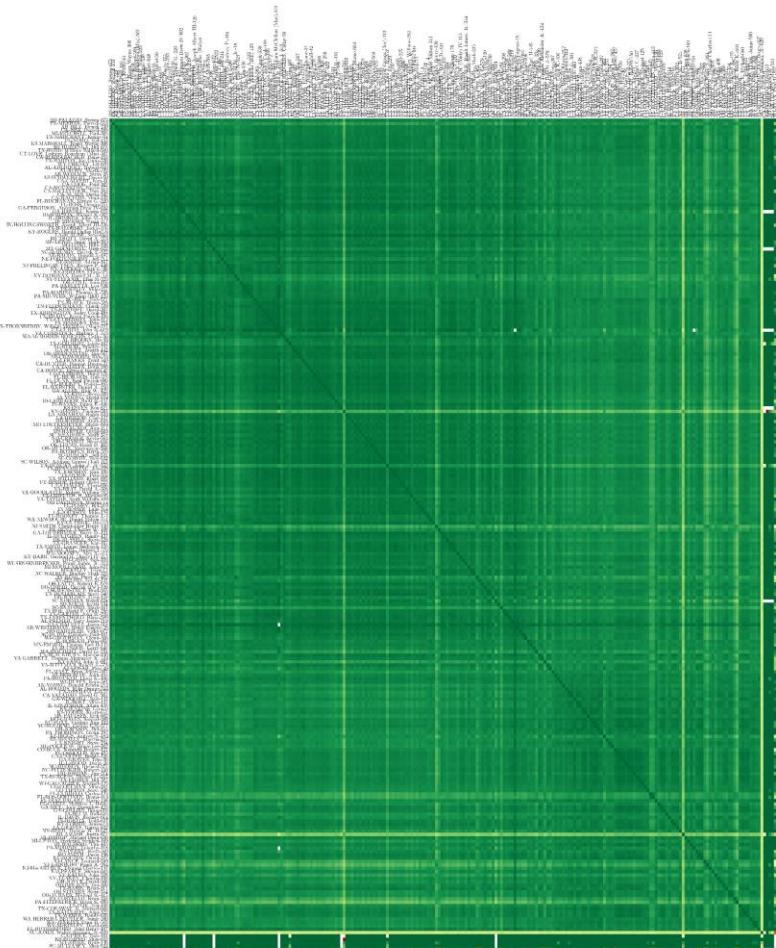


Entities (voting sessions)

<i>ide</i>	Date	Themes
<i>e</i> ₁	2016/04/16	1.05 National Budget
<i>e</i> ₂	2016/05/16	5.03 Employee benefits
<i>e</i> ₃	2016/06/16	7 Environment
<i>e</i> ₄	2016/06/16	1.20 Citizen's rights; 20.11 Gov. Adm. Issues
<i>e</i> ₅	2016/07/03	20.11 Gov. Adm. Issues
<i>e</i> ₆	2016/07/29	20.11 Gov. Adm. Issues

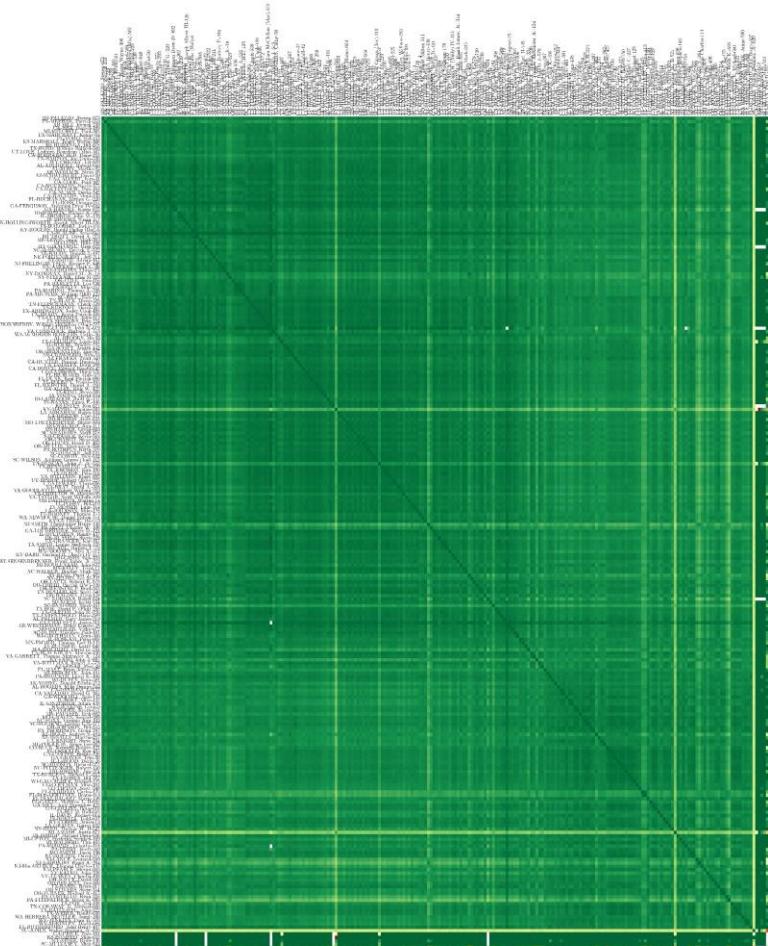
Context = Government and Administrative Issues related voting sessions

Example – in the U.S. Congress (House of Representatives)

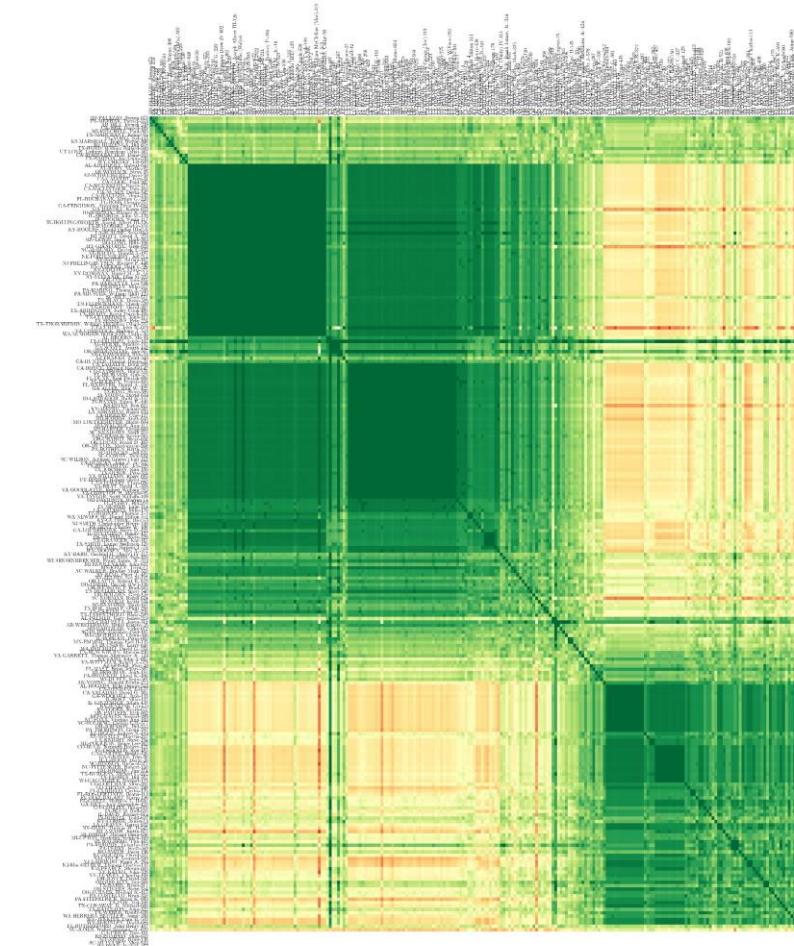


Overall **intra-agreement**
between Republicans
Group

Example – in the U.S. Congress (House of Representatives)

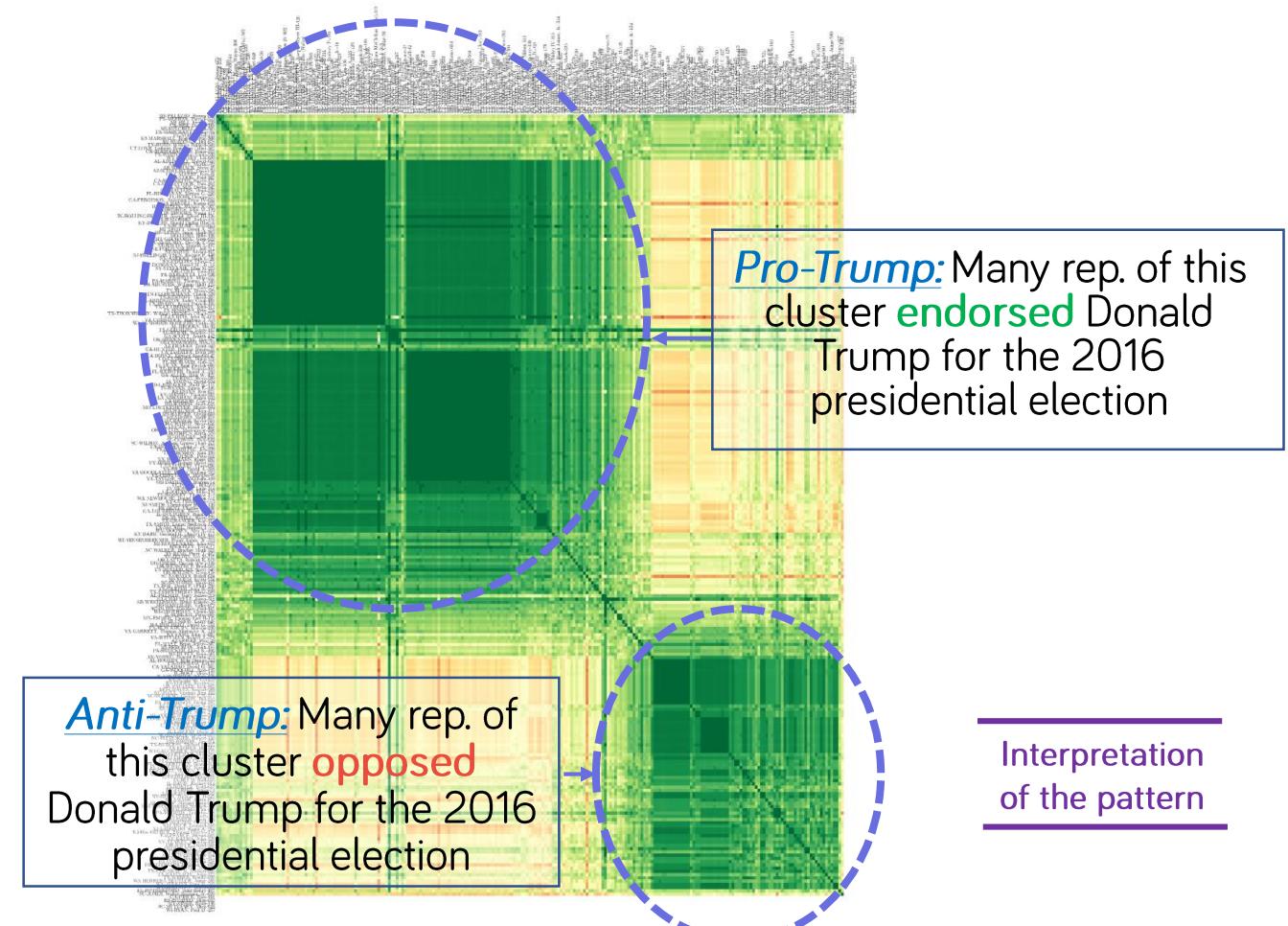
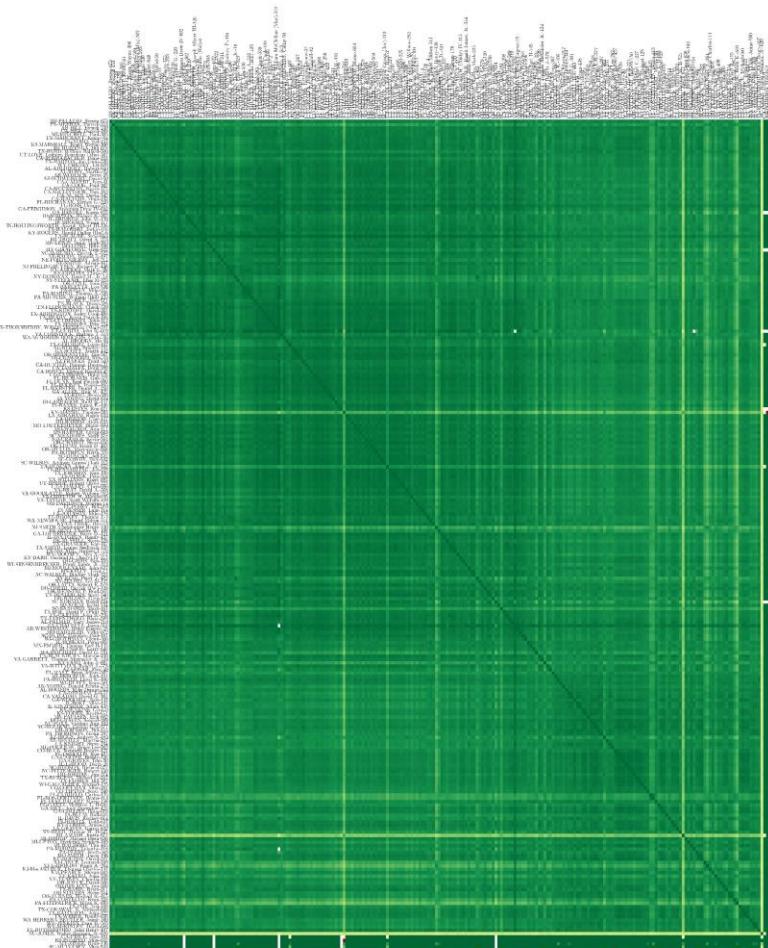


Overall **intra-agreement**
between **Republicans**
Group



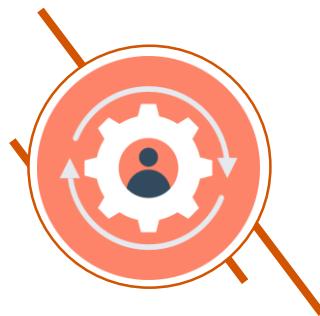
Contextual **intra-agreement** between **Republicans** in
Government and Administrative Issues related voting sessions
Context

Example – in the U.S. Congress (House of Representatives)



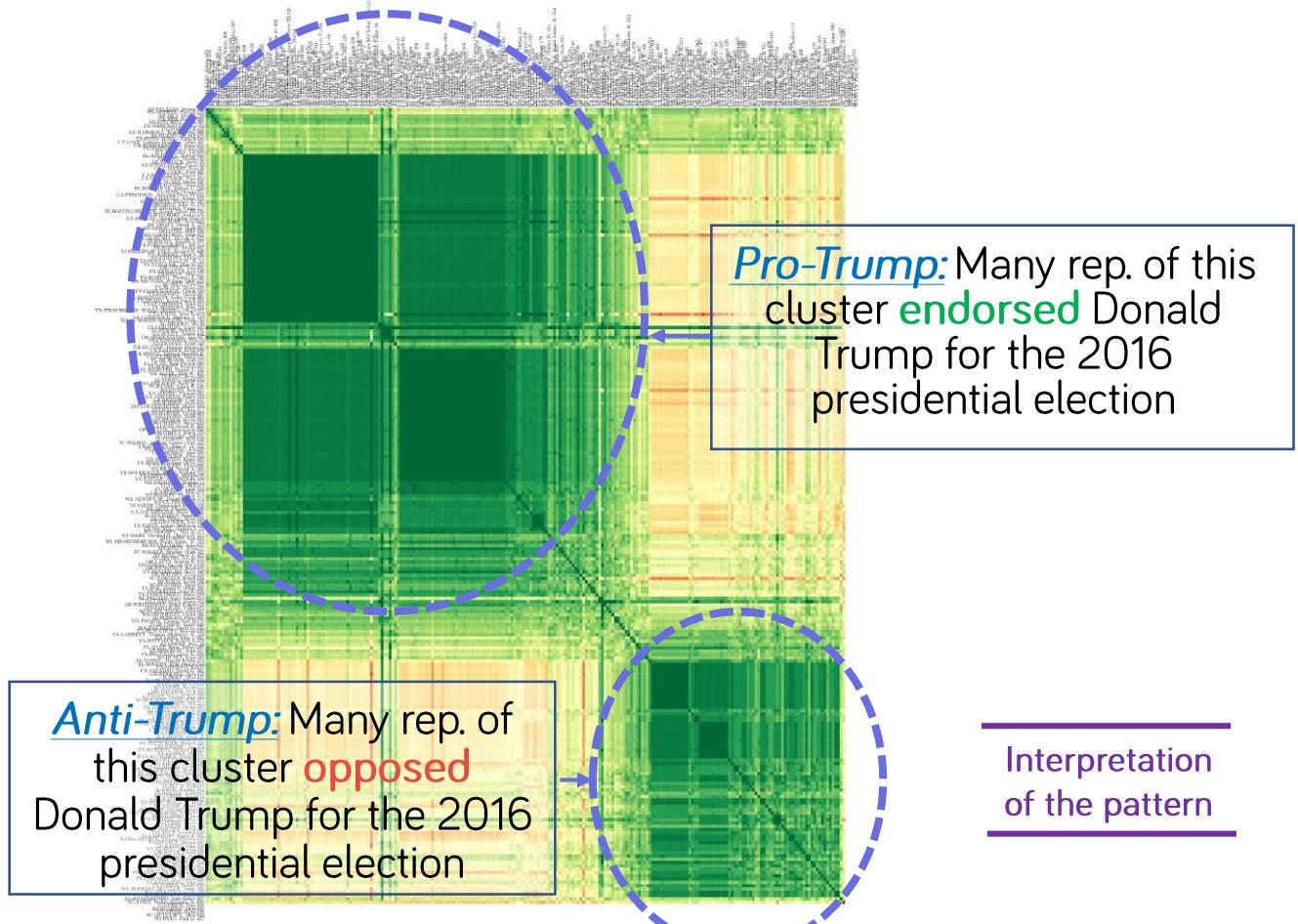
Contextual **intra-agreement** between **Republicans** in **Government and Administrative Issues** related voting sessions Context

Problem definition



Problem definition

Find exceptional
contextual intra-group
agreement patterns
(group, context)



Contextual intra-agreement between **Republicans** in
Government and Administrative Issues related voting sessions
Context

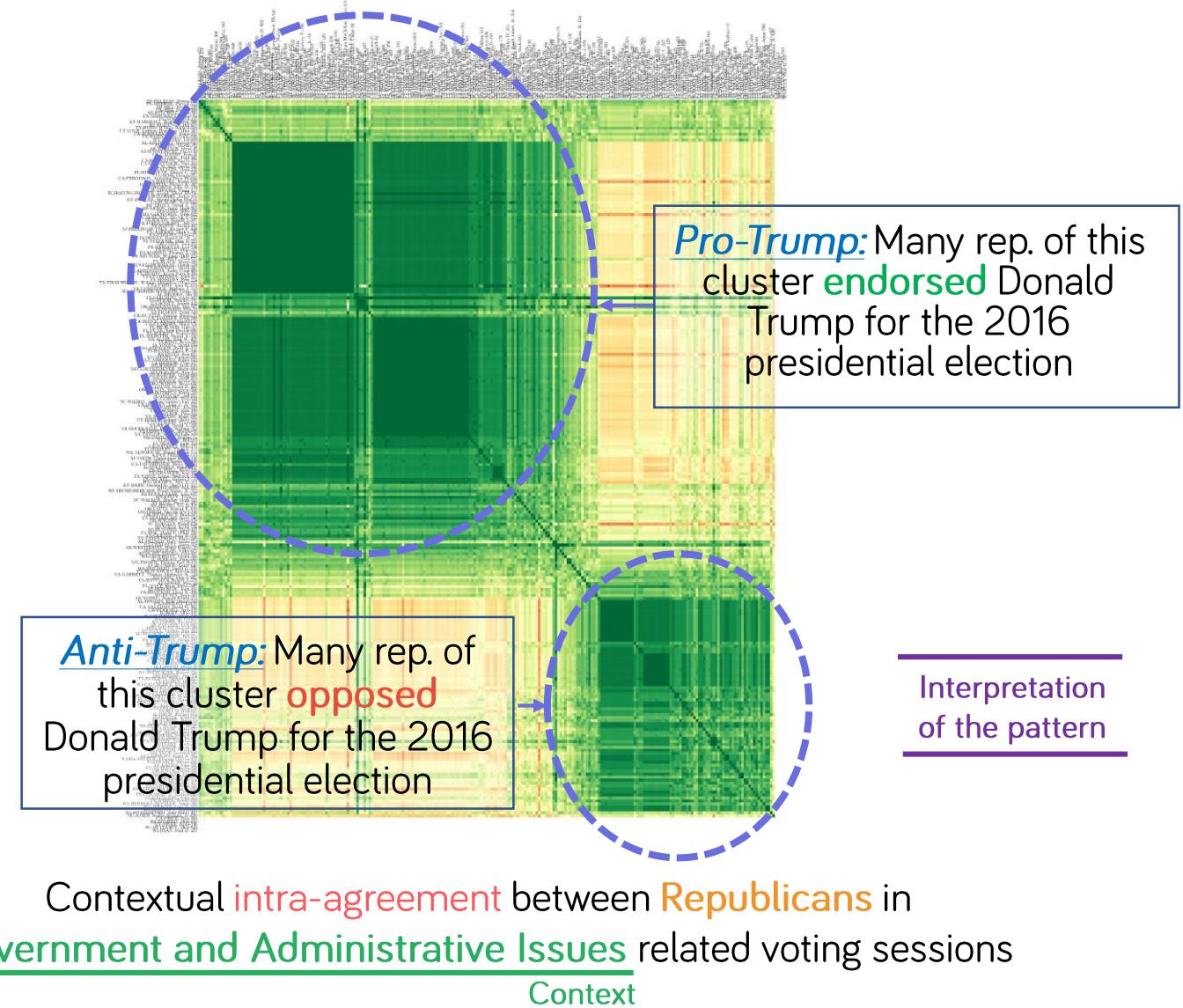
Problem definition



Problem definition

Find exceptional
contextual intra-group
agreement patterns
(group, context)

Not only in voting data
but also in other kind of
behavioral data (e.g.
rating datasets)



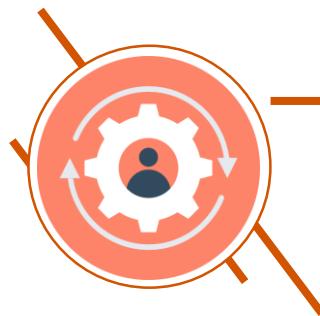


Problem definition

Find exceptional
contextual intra-group
agreement patterns
(group, context)

Not only in voting data
but also in other kind of
behavioral data (e.g.
rating datasets)

Questions



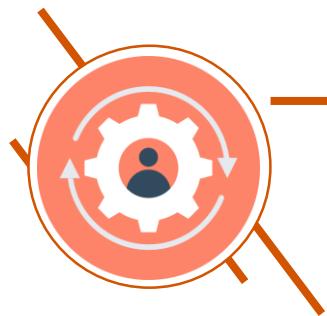
Problem definition

Find exceptional
contextual intra-group
agreement patterns
(group, context)

Not only in voting data
but also in other kind of
behavioral data (e.g.
rating datasets)

Questions

1. How to characterize **groups** and **contexts**?



Problem definition

Find exceptional
contextual intra-group
agreement patterns
(group, context)

Not only in voting data
but also in other kind of
behavioral data (e.g.
rating datasets)

Questions

1. How to characterize **groups** and **contexts**?

2. How to measure **intra-group agreement**?

- applies to **any number** of individuals
- handles **various domains** of outcomes (categorical, numerical, ordinal ...)
- handles **missing values**
- corrects for the agreement **expected by chance**



Problem definition

Find exceptional
contextual intra-group
agreement patterns
(group, context)

Not only in voting data
but also in other kind of
behavioral data (e.g.
rating datasets)

Questions

1. How to characterize **groups** and **contexts**?
2. How to measure **intra-group agreement**?
 - applies to **any number** of individuals
 - handles **various domains** of outcomes (categorical, numerical, ordinal ...)
 - handles **missing values**
 - corrects for the agreement **expected by chance**
3. How to evaluate the **exceptionality** (interestingness) of a pattern?



Problem definition

Find exceptional
contextual intra-group
agreement patterns
(group, context)

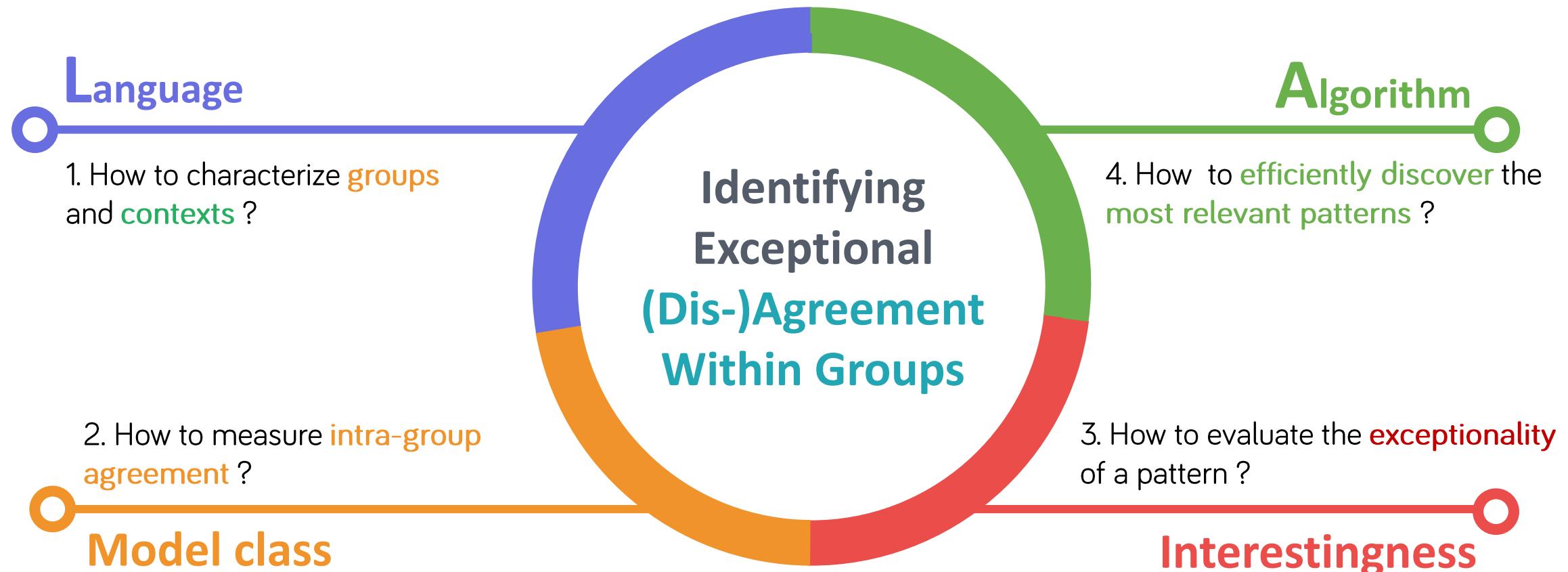
Not only in voting data
but also in other kind of
behavioral data (e.g.
rating datasets)

Questions

1. How to characterize **groups** and **contexts**?
2. How to measure **intra-group agreement**?
 - applies to **any number** of individuals
 - handles **various domains** of outcomes (categorical, numerical, ordinal ...)
 - handles **missing values**
 - corrects for the agreement **expected by chance**
3. How to evaluate the **exceptionality** (interestingness) of a pattern?
4. How to **efficiently discover** the **most relevant patterns**?



Leman, D., Feelders, A., & Knobbe, A. Exceptional model mining. In ECML/PKDD 2008



* Leman, D., Feelders, A., & Knobbe, A. Exceptional model mining. In ECML/PKDD 2008

Language
Algorithm
**Identifying
Exceptional
(Dis-)Agreement
Within Groups**
Interestingness
Model class



DEvIANT IN A NUTSHELL

Language

Heterogeneous Attribute-Value Data (Categorical, Numerical, Hierarchical):

Conjunctions of conditions on single attributes



Individuals (Parliamentarians)				Entities (voting sessions)		
<i>idi</i>	<i>Party</i>	<i>State</i>	<i>Age</i>	<i>ide</i>	<i>Date</i>	<i>Themes</i>
i_1	Democrats	New York	32	e_1	2016/04/16	1.05 National Budget
i_2	Democrats	Florida	41	e_2	2016/05/16	5.03 Employee benefits
i_3	Republicans	Texas	45	e_3	2016/06/16	7 Environment
i_4	Republicans	Chicago	39	e_4	2016/06/16	1.20 Citizen's rights; 20.11 Gov. Adm. Issues
i_5	Republicans	Washington	28	e_5	2016/07/03	20.11 Gov. Adm. Issues
i_6	Republicans	Illinois	51	e_6	2016/07/29	20.11 Gov. Adm. Issues

Party = Republicans **Context = Government and Administrative Issues related voting sessions**

Language

Heterogeneous Attribute-Value Data (Categorical, Numerical, Hierarchical):

Conjunctions of conditions on single attributes



Individuals (Parliamentarians)				Entities (voting sessions)		
<i>idi</i>	<i>Party</i>	<i>State</i>	<i>Age</i>	<i>ide</i>	<i>Date</i>	<i>Themes</i>
<i>i</i> ₁	Democrats	New York	32	<i>e</i> ₁	2016/04/16	1.05 National Budget
<i>i</i> ₂	Democrats	Florida	41	<i>e</i> ₂	2016/05/16	5.03 Employee benefits
<i>i</i> ₃	Republicans	Texas	45	<i>e</i> ₃	2016/06/16	7 Environment
<i>i</i> ₄	Republicans	Chicago	39	<i>e</i> ₄	2016/06/16	1.20 Citizen's rights; 20.11 Gov. Adm. Issues
<i>i</i> ₅	Republicans	Washington	28	<i>e</i> ₅	2016/07/03	20.11 Gov. Adm. Issues
<i>i</i> ₆	Republicans	Illinois	51	<i>e</i> ₆	2016/07/29	20.11 Gov. Adm. Issues

Party = Republicans

Context = Government and Administrative Issues related voting sessions

Pattern = (group, context) =
(Party = Republicans , Themes ⊇ {20.11 Gov. Adm. Issues})

Language

Heterogeneous Attribute-Value Data (Categorical, Numerical, Hierarchical):

Conjunctions of conditions on single attributes



Individuals (Parliamentarians)				Entities (voting sessions)		
<i>idi</i>	<i>Party</i>	<i>State</i>	<i>Age</i>	<i>ide</i>	<i>Date</i>	<i>Themes</i>
i_1	Democrats	New York	32	e_1	2016/04/16	1.05 National Budget
i_2	Democrats	Florida	41		2016/05/16	5.03 Employee benefits
i_3	Republicans	Texas	45		2016/06/16	7 Environment
i_4	Republicans	Chicago	39		2016/06/16	1.20 Citizen's rights; 20.11 Gov. Adm. Issues
i_5	Republicans	Washington	28		2016/07/03	20.11 Gov. Adm. Issues
i_6	Republicans	Illinois	51		2016/07/29	20.11 Gov. Adm. Issues

Party = Republicans **Context = Government and Administrative Issues related voting sessions**

Pattern = (group, context) =
(Party = Republicans , Themes \supseteq {20.11 Gov. Adm. Issues})

- Categorical: Equality Test
- Numerical: Interval Memberships
- Hierarchical: Subset Memberships

Model Class

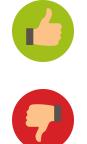
How to measure **intra-group agreement** ?



Selected group:
Republicans

BD	e_1	e_2	e_3	e_4	e_5	e_6
i_3		👎			👍	👎
i_4	👍		👍	👎	👍	
i_5	👍	👎		👍	👎	
i_6	👍		👍			👎

8	👍
6	👎



- It is applicable to **any number** of individuals
- It handles **various domain** of outcomes
- It handles **missing values**
- It corrects for the agreement **expected by chance**

* Hayes, A. F., & Krippendorff, K. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1), 77-89, 2007

Model Class

How to measure **intra-group agreement** ?

$$A(c) = 1 - \frac{D_{obs}(c)}{D_{exp}}$$



Hayes, A. F., & Krippendorff, K. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1), 77-89, 2007

Selected group:
Republicans

<i>BD</i>	<i>e</i> ₁	<i>e</i> ₂	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆
<i>i</i> ₃		👎			👍	👎
<i>i</i> ₄	👍		👍	👎	👍	
<i>i</i> ₅	👍	👎		👍	👎	
<i>i</i> ₆	👍		👍			👎

8	👍
6	👎

Krippendorff's Alpha

- It is applicable to **any number** of individuals
- It handles **various domain** of outcomes
- It handles **missing values**
- It corrects for the agreement **expected by chance**

Model Class

How to measure **intra-group agreement** ?

$$A(c) = 1 - \frac{D_{obs}(c)}{D_{exp}}$$



Hayes, A. F., & Krippendorff, K. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1), 77-89, 2007

Selected group:
Republicans

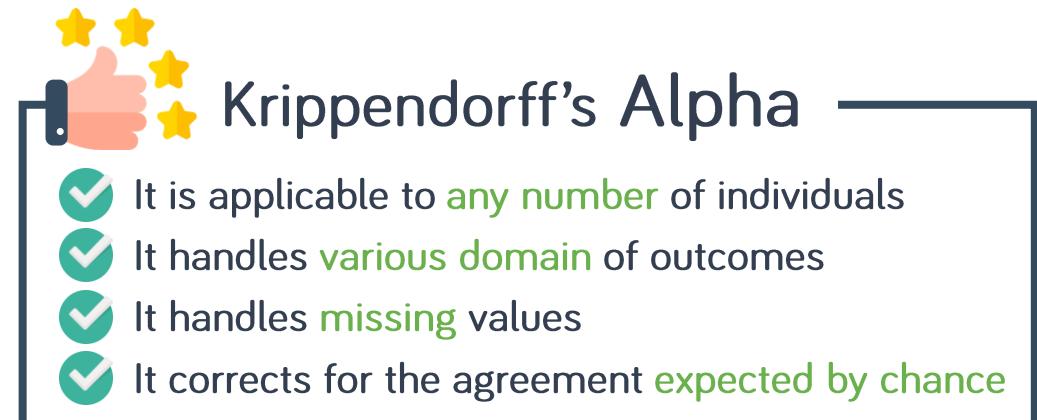
<i>BD</i>	<i>e</i> ₁	<i>e</i> ₂	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆
<i>i</i> ₃						
<i>i</i> ₄						
<i>i</i> ₅						
<i>i</i> ₆						

8	
6	

<i>m</i> _e	3	2	2	2	3	2
<i>v</i> _e	1	1	1	-0,89	-0,26	1

Alpha is a **weighted arithmetic mean**

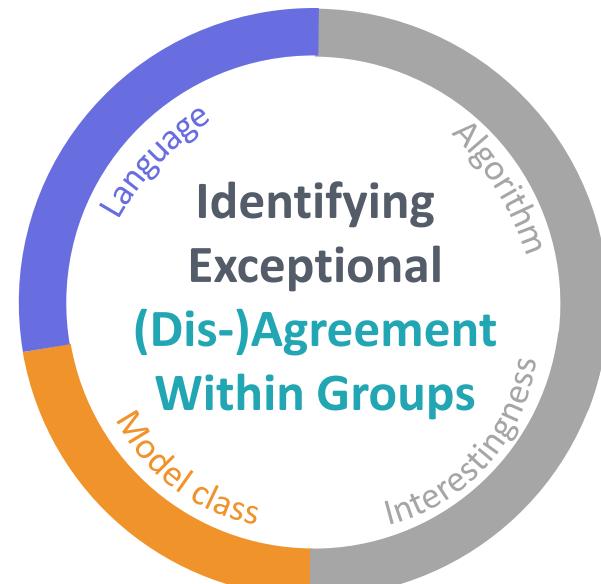
$$A(c) = \frac{\sum_{e \in G_E^c} m_e v_e}{\sum_{e \in G_E^c} m_e}$$



Model Class

How to measure **intra-group agreement** ?

$$A(c) = 1 - \frac{D_{obs}(c)}{D_{exp}}$$



Hayes, A. F., & Krippendorff, K. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1), 77-89, 2007

Selected group:
Republicans

<i>BD</i>	<i>e</i> ₁	<i>e</i> ₂	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆
<i>i</i> ₃						
<i>i</i> ₄						
<i>i</i> ₅						
<i>i</i> ₆						

8	
6	

<i>m</i> _e	3	2	2	2	3	2
<i>v</i> _e	1	1	1	-0,89	-0,26	1

Alpha is a **weighted arithmetic mean**

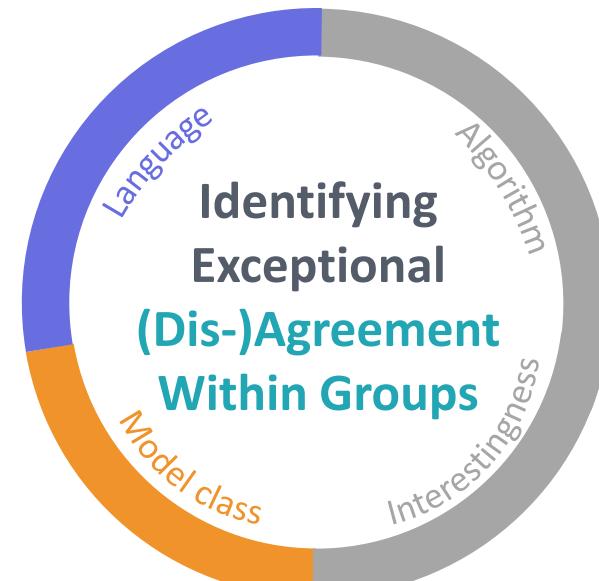
$$A(c) = \frac{\sum_{e \in G_E^c} m_e v_e}{\sum_{e \in G_E^c} m_e}$$



Model Class

How to measure **intra-group agreement** ?

$$A(c) = 1 - \frac{D_{obs}(c)}{D_{exp}}$$



Hayes, A. F., & Krippendorff, K. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1), 77-89, 2007

Selected group:
Republicans

<i>BD</i>	<i>e</i> ₁	<i>e</i> ₂	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆
<i>i</i> ₃						
<i>i</i> ₄						
<i>i</i> ₅						
<i>i</i> ₆						

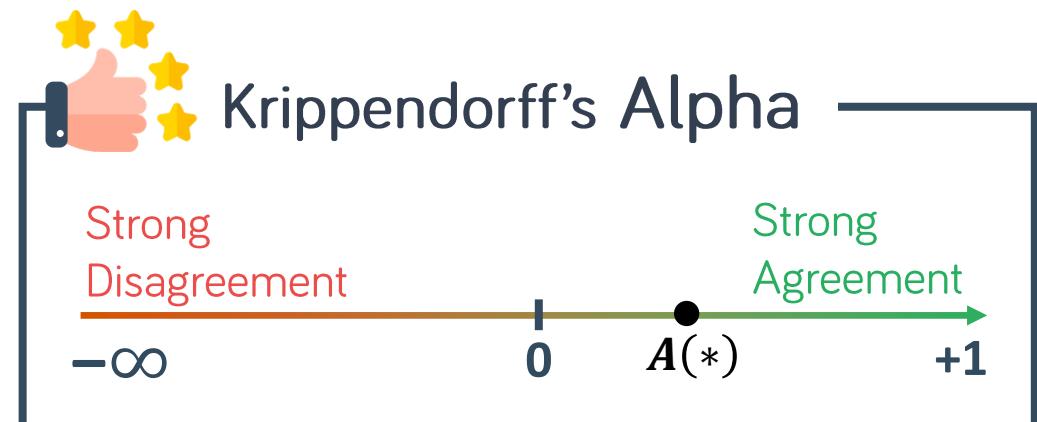
8	
6	

<i>m</i> _e	3	2	2	2	3	2
<i>v</i> _e	1	1	1	-0,89	-0,26	1

$$A(*) = 0,46$$

Alpha is a **weighted arithmetic mean**

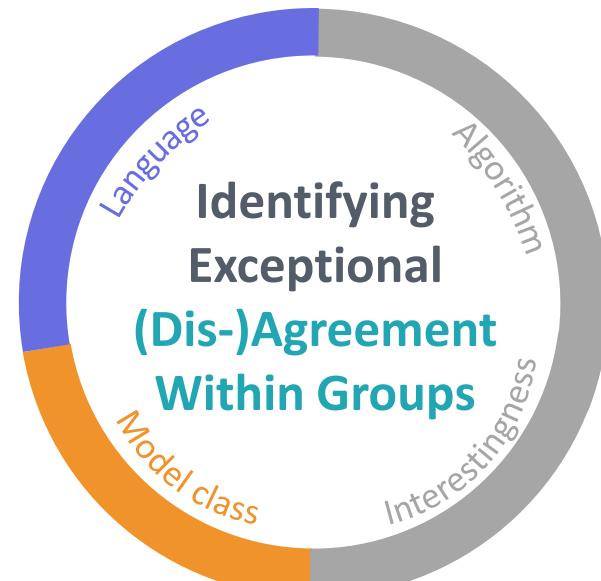
$$A(c) = \frac{\sum_{e \in G_E^c} m_e v_e}{\sum_{e \in G_E^c} m_e}$$



Model Class

How to measure **intra-group agreement**?

$$A(c) = 1 - \frac{D_{obs}(c)}{D_{exp}}$$



Hayes, A. F., & Krippendorff, K. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1), 77-89, 2007

Selected group:
Republicans

<i>BD</i>	<i>e</i> ₁	<i>e</i> ₂	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆
<i>i</i> ₃						
<i>i</i> ₄						
<i>i</i> ₅						
<i>i</i> ₆						
<i>m</i> _e	3	2	2	2	3	2
<i>v</i> _e	1	1	1	-0,89	-0,26	1

**Context = Gov.
Admin. Issues**

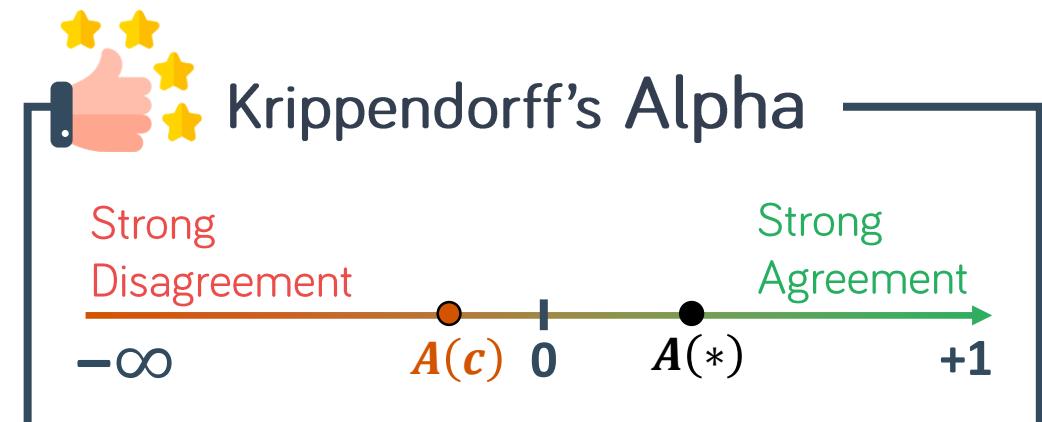
8	
6	

$$A(*) = 0,46$$

$$A(c) = -0,08$$

Alpha is a **weighted arithmetic mean**

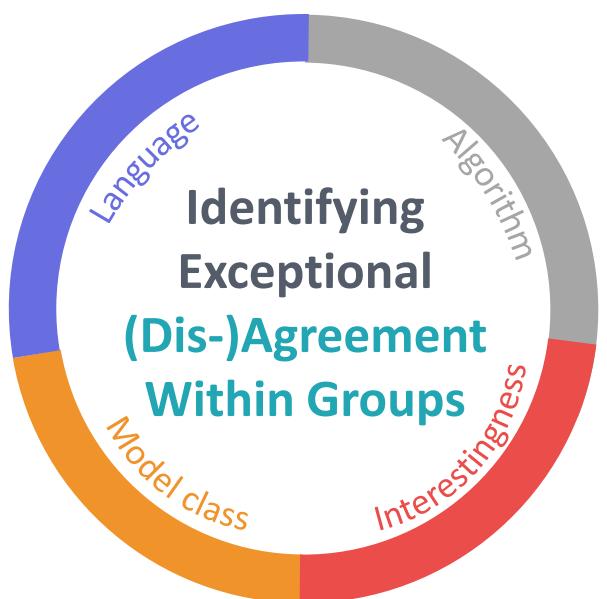
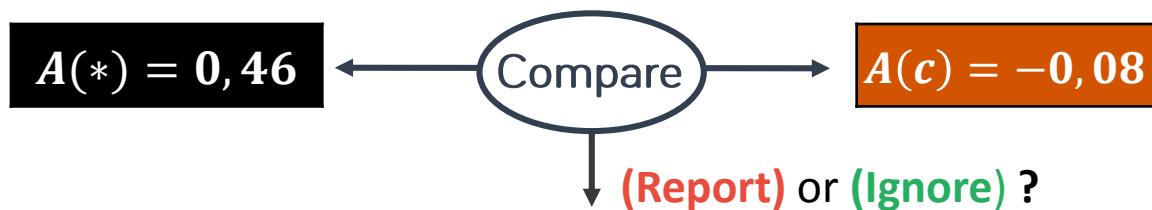
$$A(c) = \frac{\sum_{e \in G_E^c} m_e v_e}{\sum_{e \in G_E^c} m_e}$$



Pattern = (Country = Germany ,Themes ⊇ {20.11 Gov. Adm. Issues})

Interestingness

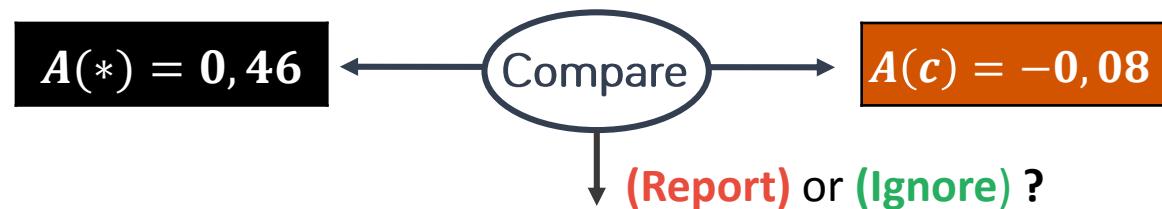
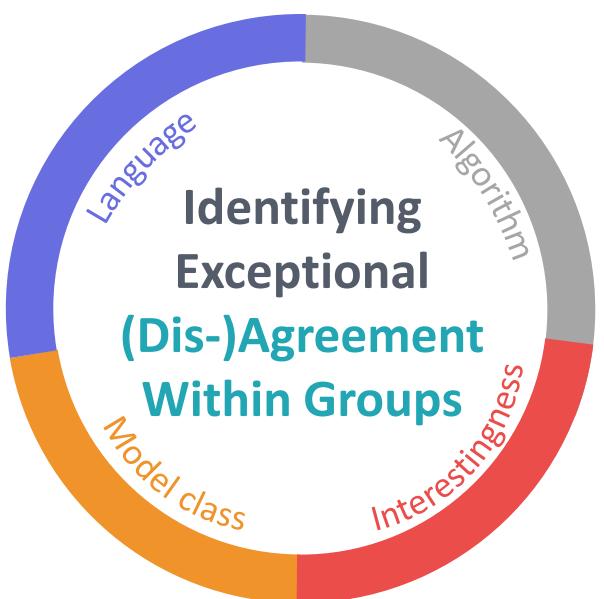
How to evaluate the **exceptionality**
(interestingness) of a pattern ?



Pattern = (Country = Germany ,Themes ⊇ {20.11 Gov. Adm. Issues})

Interestingness

How to evaluate the **exceptionality** (interestingness) of a pattern ?



Questions to tackle in case of simple comparison function (difference)

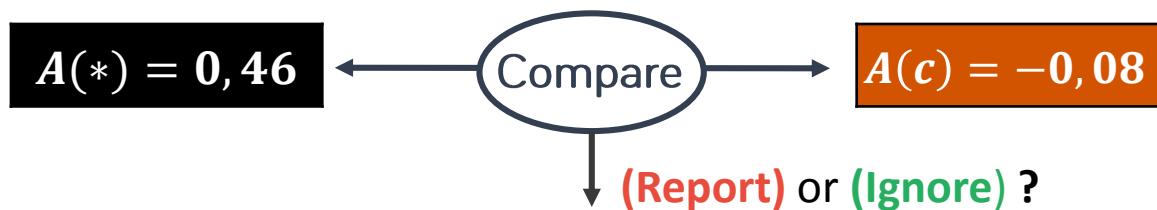


- What **minimum threshold** to choose?
- How to handle **exceptionality of intra-group agreement** at **different scales** of **subgroups size** ?
- How to avoid reporting spurious findings (**statistical significance**) ?

Pattern = (Country = Germany ,Themes ⊇ {20.11 Gov. Adm. Issues})

Interestingness

How to evaluate the **exceptionality** (interestingness) of a pattern ?



Questions to tackle in case of simple comparison function (difference)

- What **minimum threshold** to choose?
- How to handle **exceptionality of intra-group agreement** at **different scales** of **subgroups size** ?
- How to avoid reporting spurious findings (**statistical significance**) ?



Solution

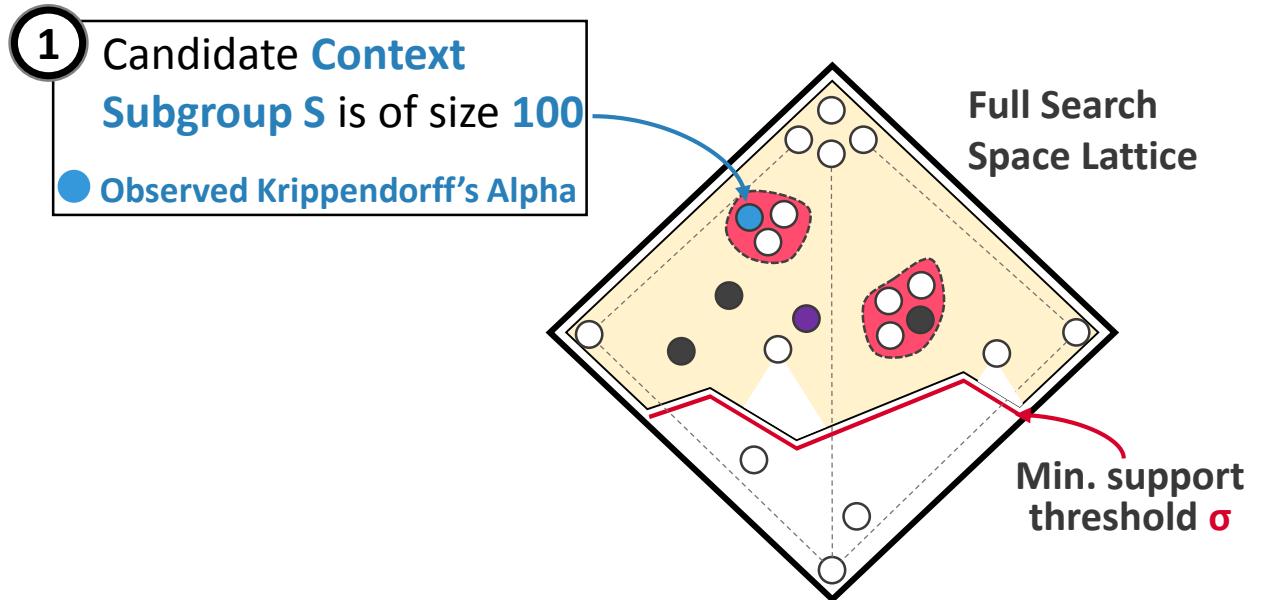
Compute the **p-value** of the **observed contextual intra-group agreement** $A(c)$ using the **DFD*** - Distribution of False Discoveries.

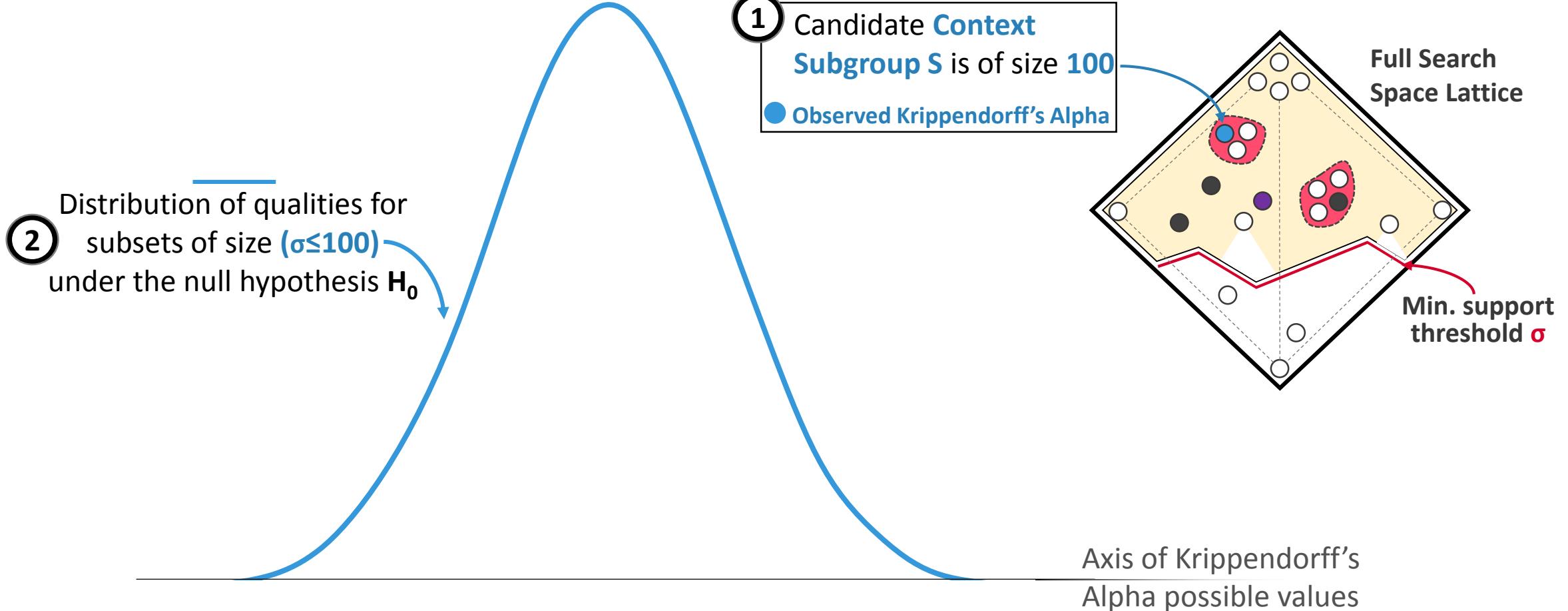


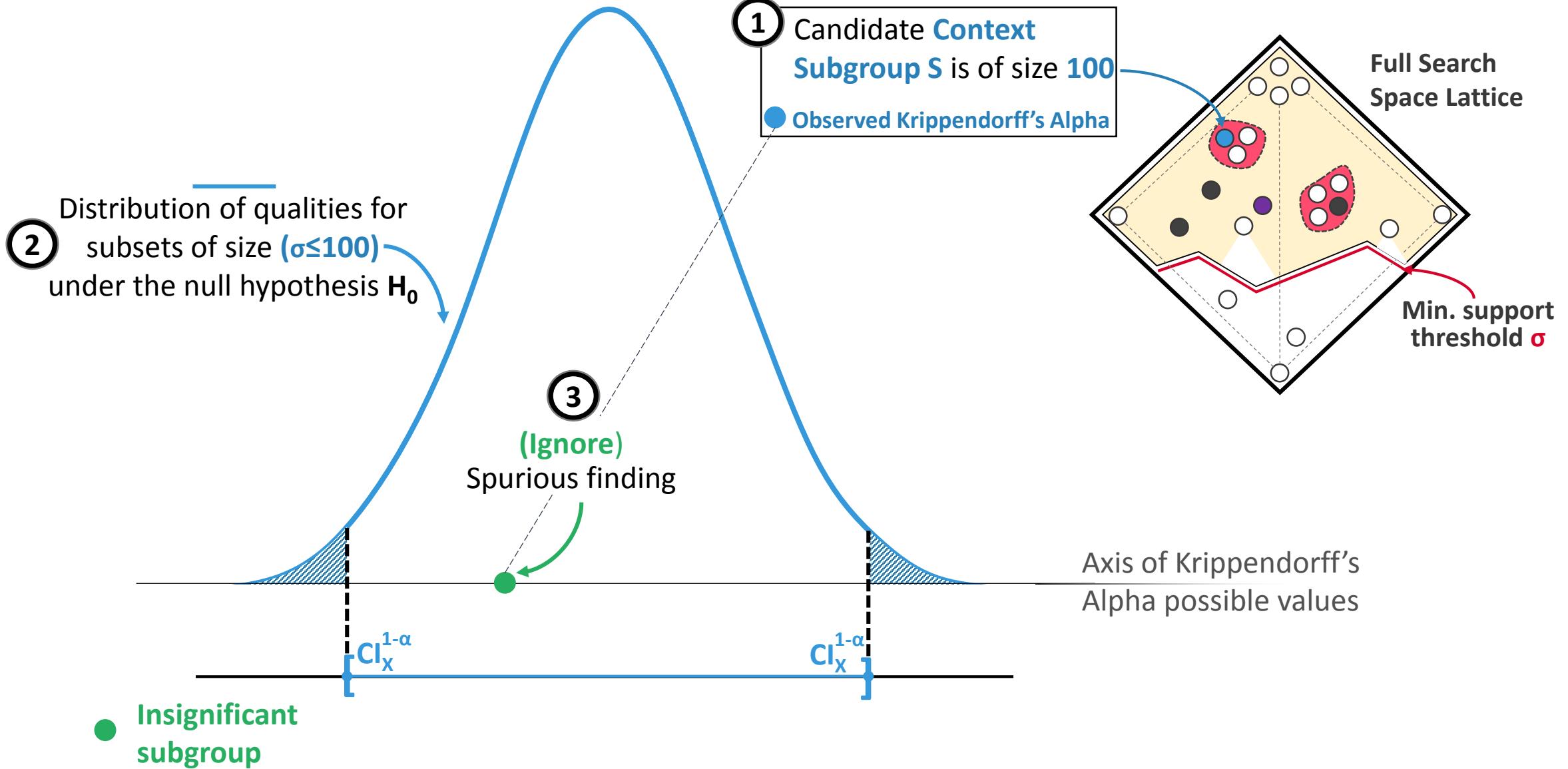
Duivesteijn, W., & Knobbe, A. Exploiting False Discoveries-Statistical Validation of Patterns and Quality Measures in Subgroup Discovery. In ICDM 2011

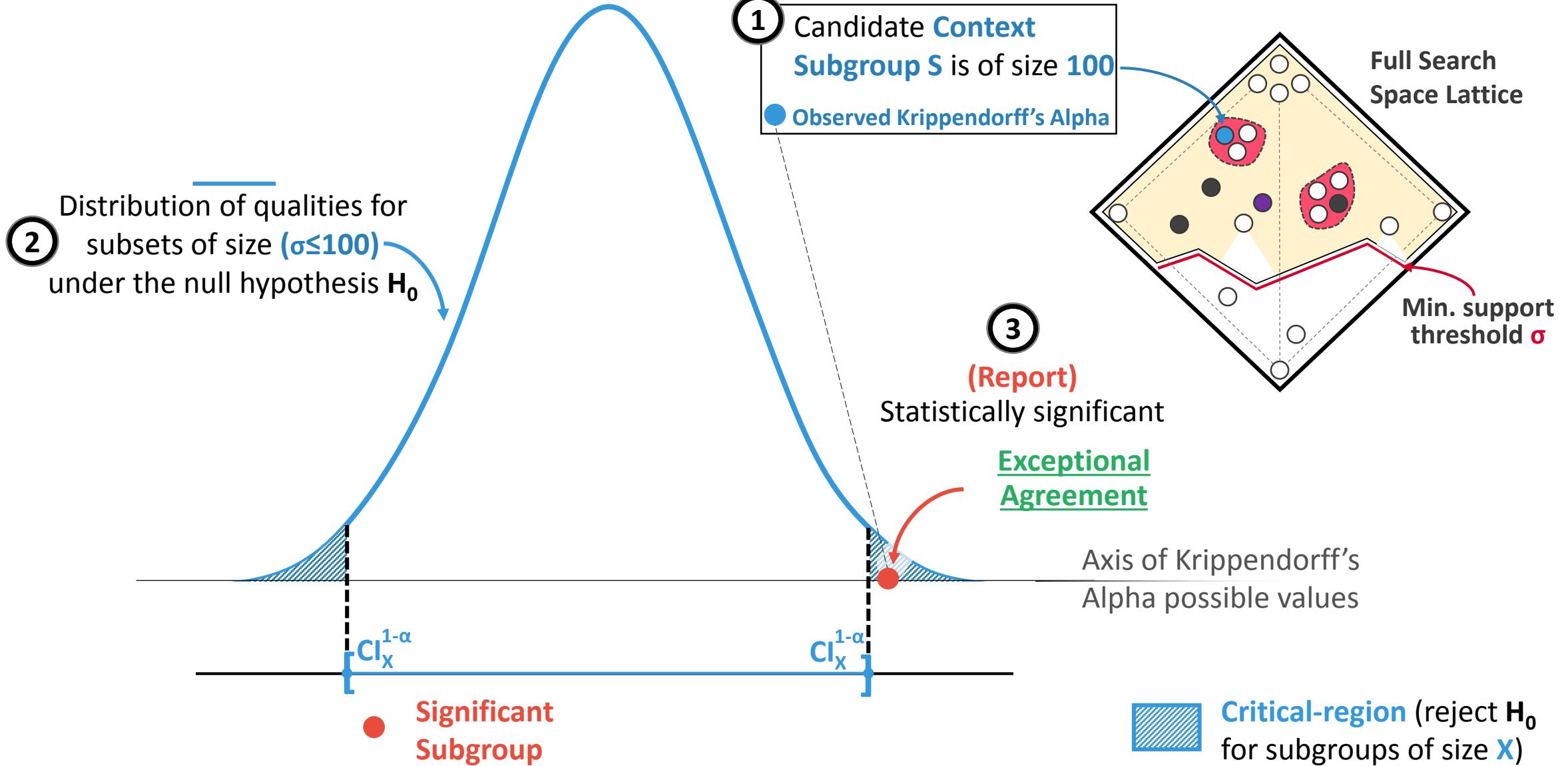


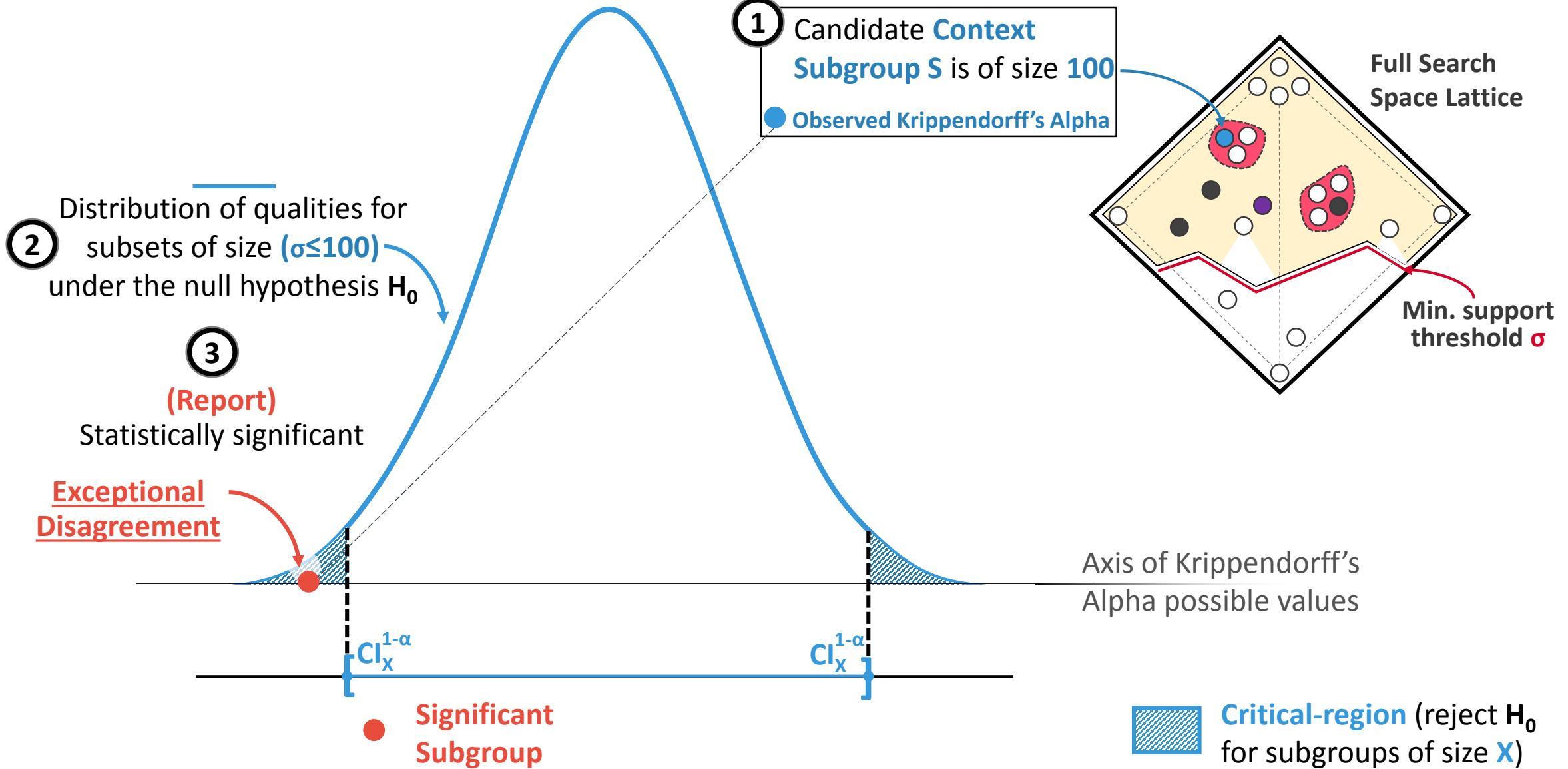
Lemmerich, F., Becker, M., Singer, P., Helic, D., Hotho, A., & Strohmaier, M. Mining subgroups with exceptional transition behavior. In KDD 2016.



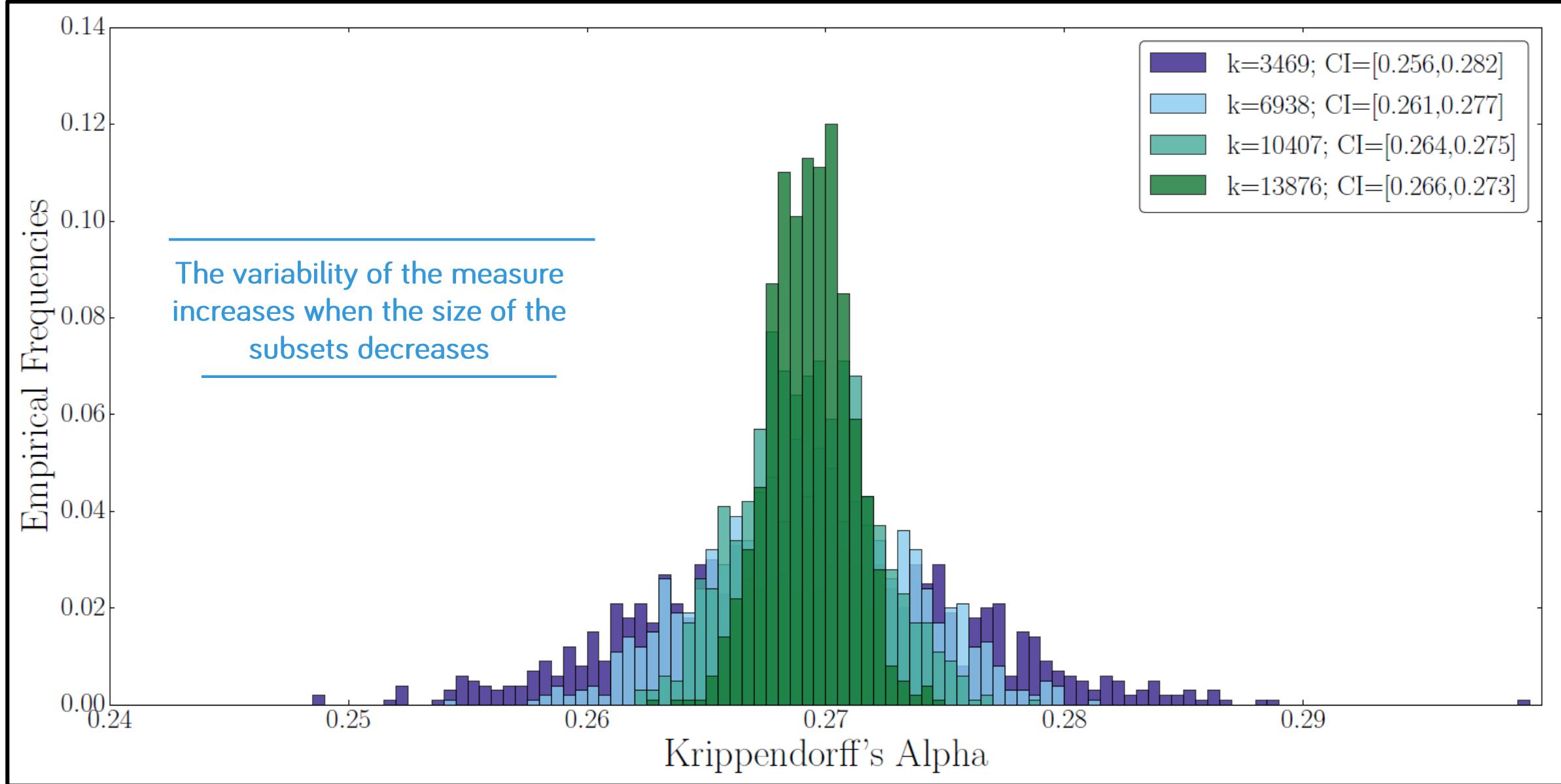


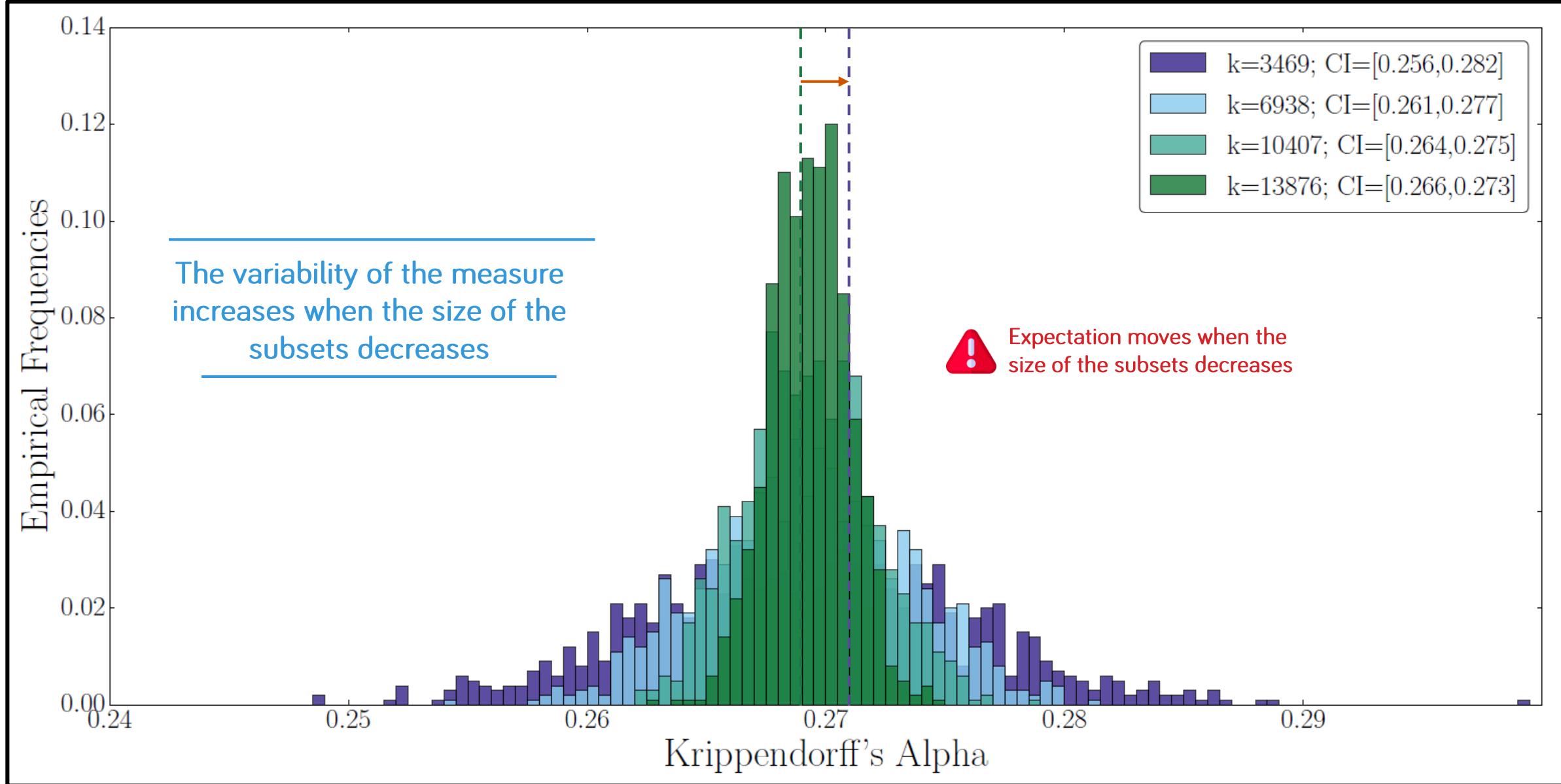






Critical-region (reject H_0 for subgroups of size X)



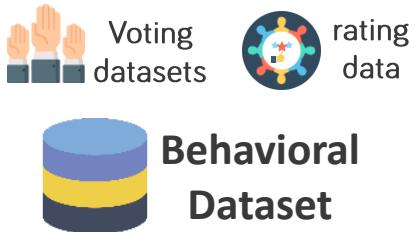


Algorithm

How to efficiently discover the most relevant patterns ?

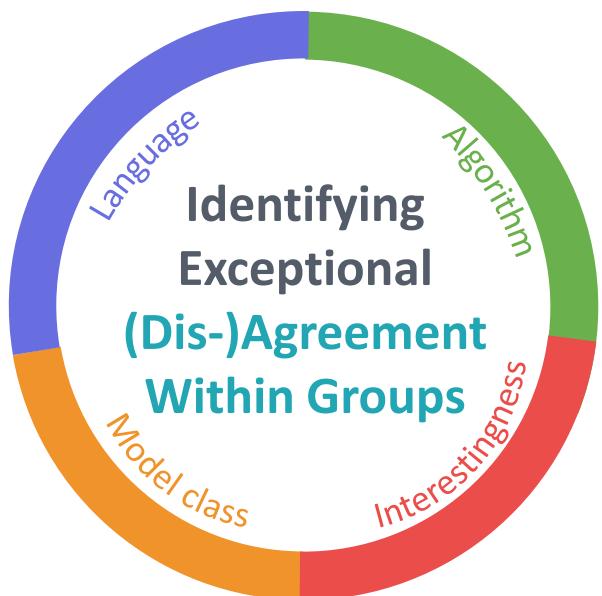


DEvIANT | Discovering Exceptional Intra-group AgreemeNt paTterns



Algorithm

How to efficiently discover the most relevant patterns ?



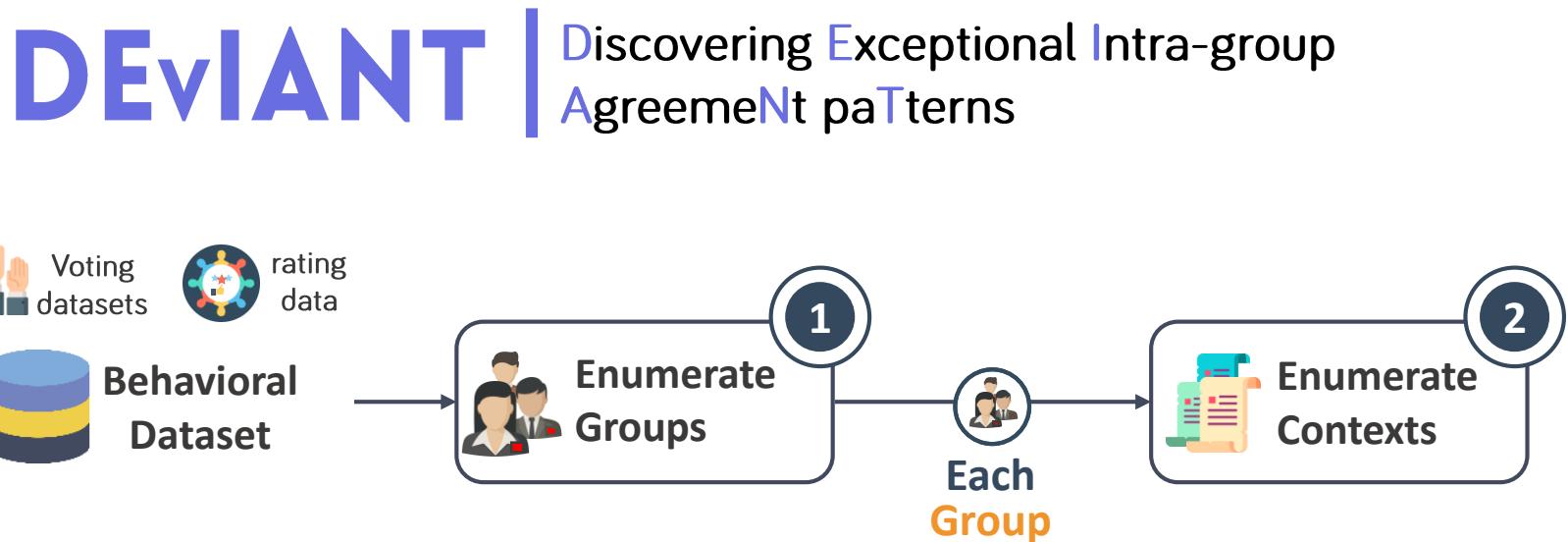
DEvIANT

Discovering Exceptional Intra-group AgreemeNt paTterns



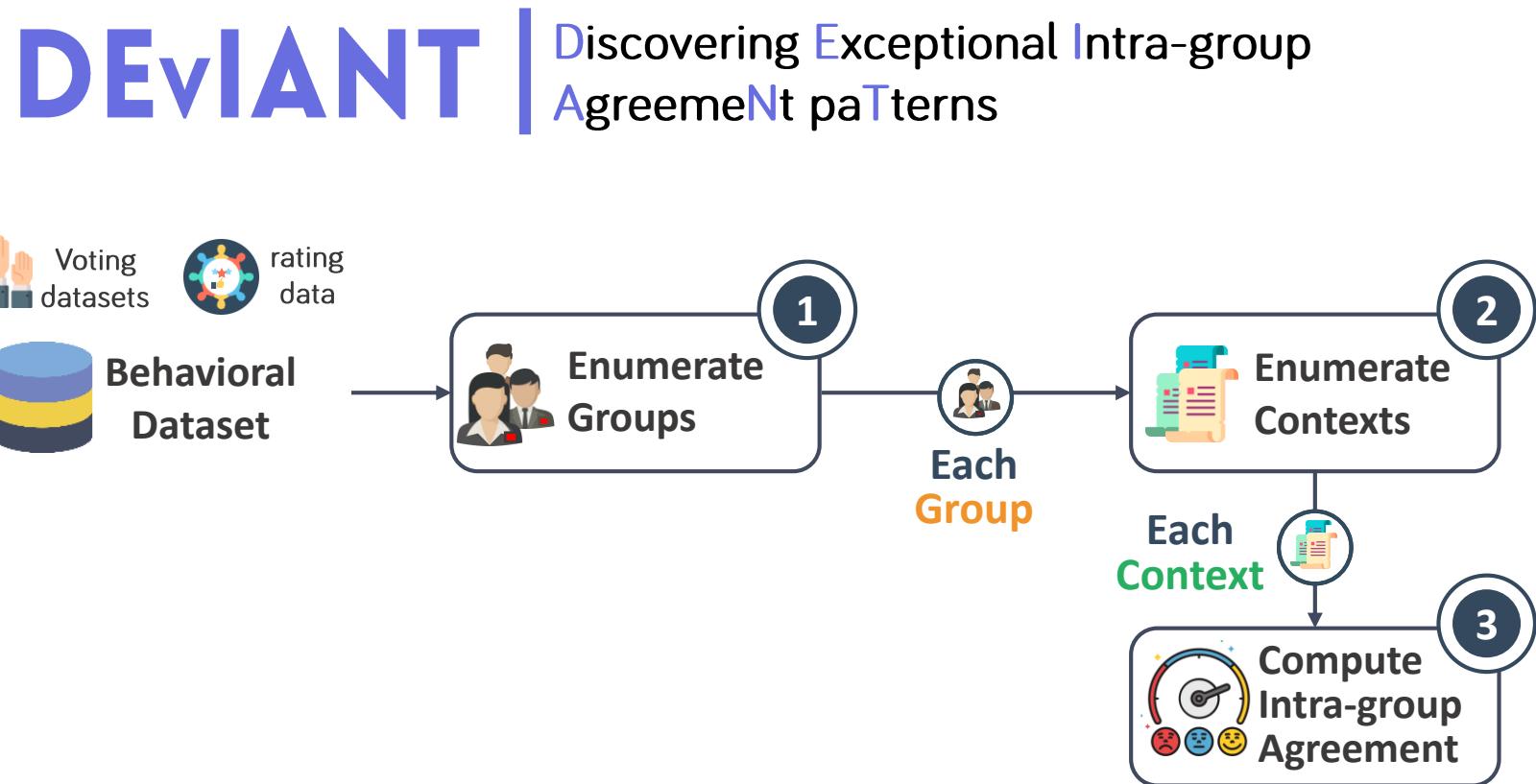
Algorithm

How to efficiently discover the most relevant patterns ?



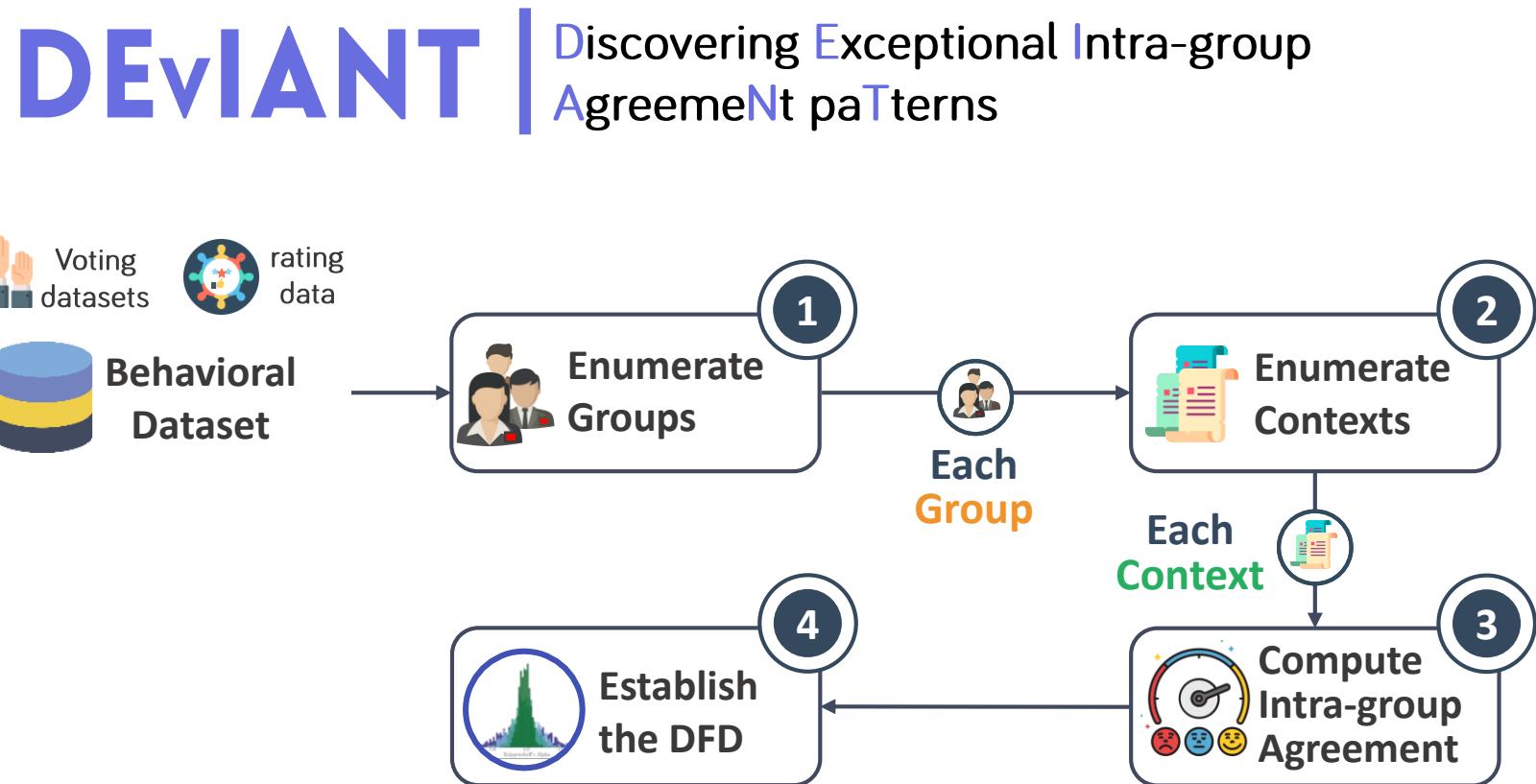
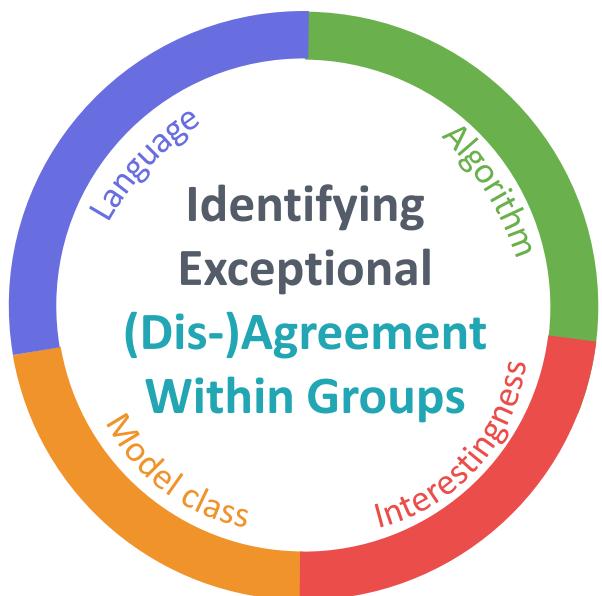
Algorithm

How to efficiently discover the most relevant patterns ?



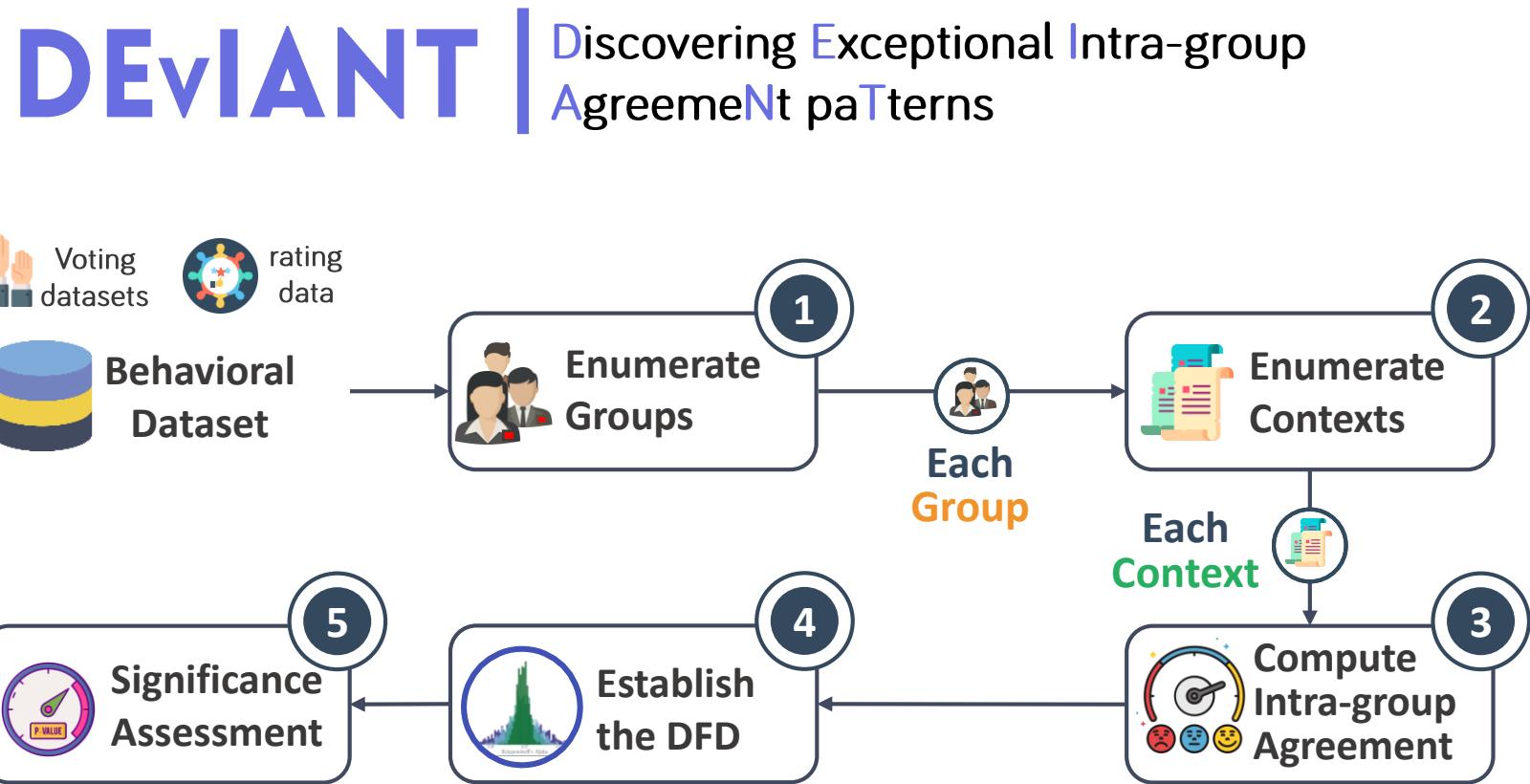
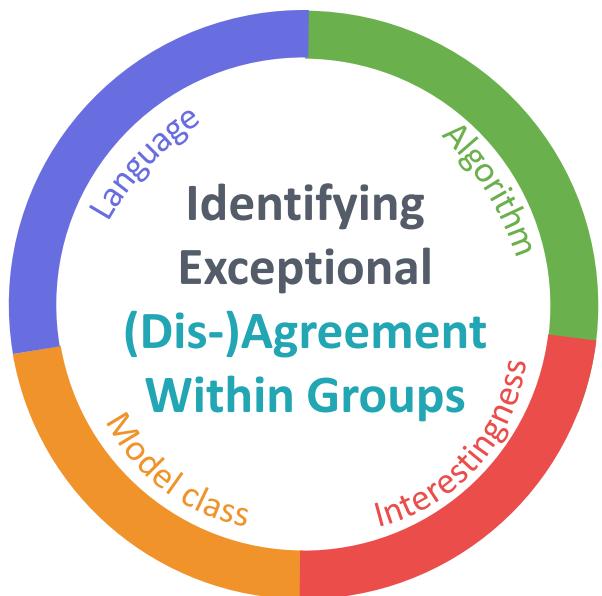
Algorithm

How to efficiently discover the most relevant patterns ?



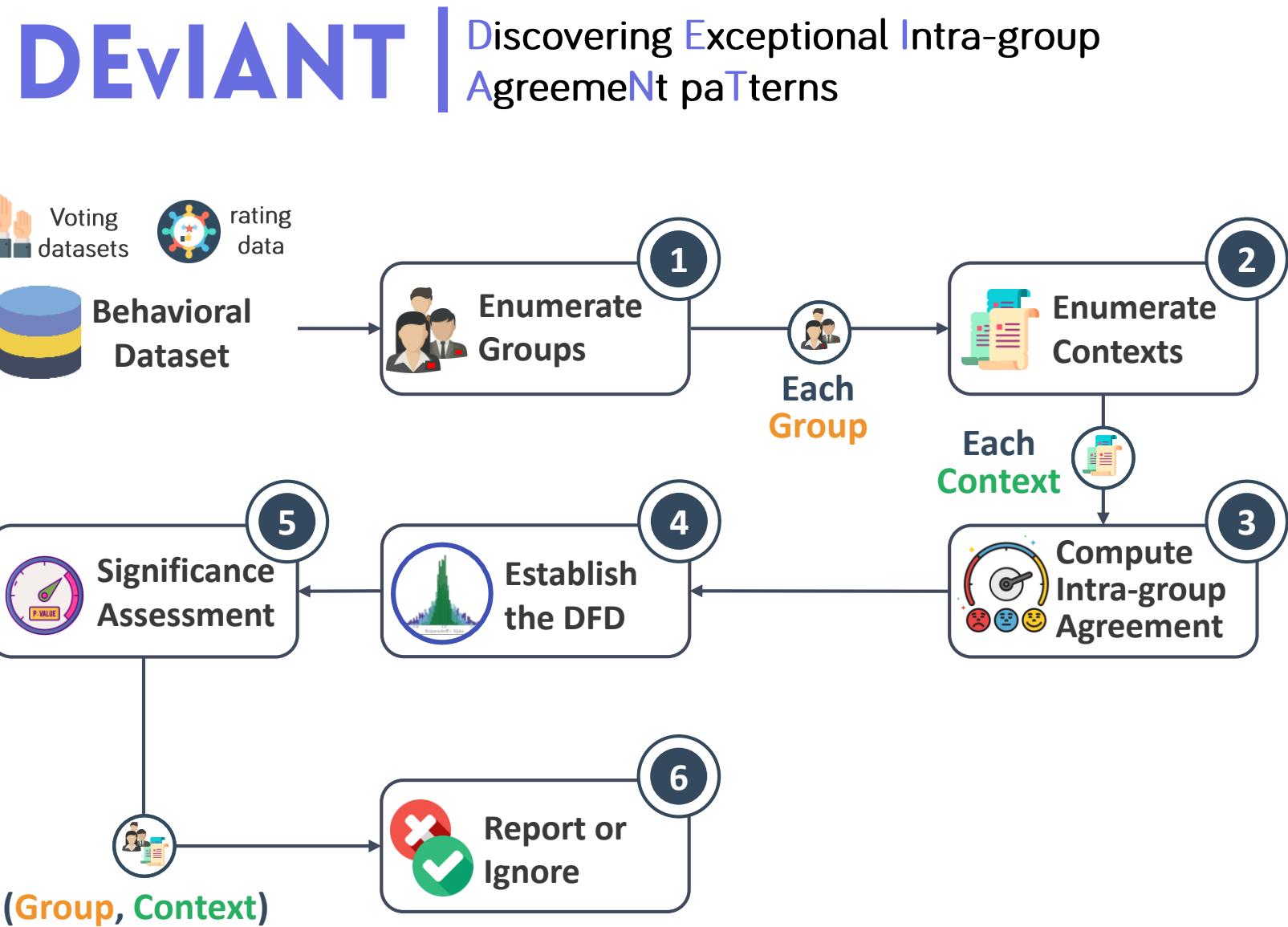
Algorithm

How to efficiently discover the most relevant patterns ?



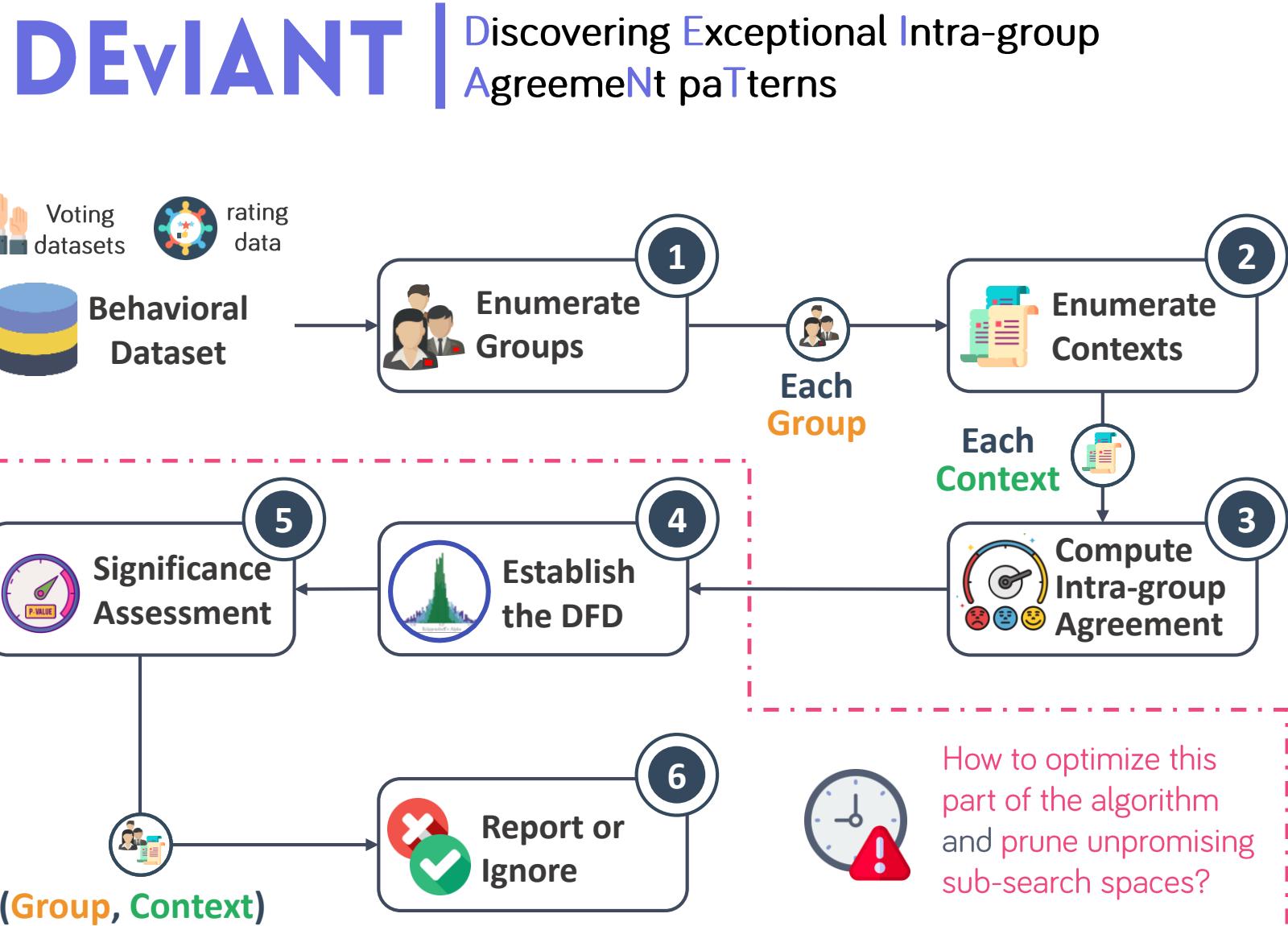
Algorithm

How to efficiently discover the most relevant patterns ?



Algorithm

How to efficiently discover the most relevant patterns ?



Algorithm

How to efficiently discover the most relevant patterns ?

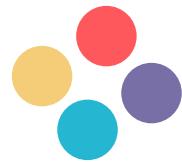
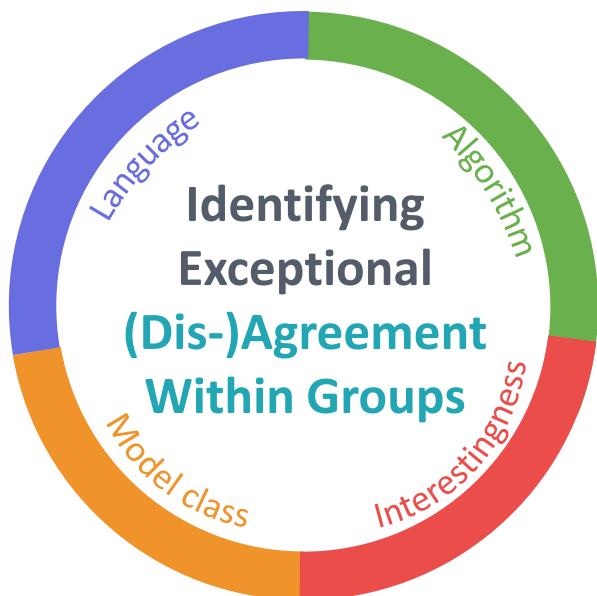


How efficient is DEvIANT?

- Compute an analytic approximate confidence interval rather than an empirical one (Linear to the size of input)

Algorithm

How to efficiently discover the most relevant patterns ?



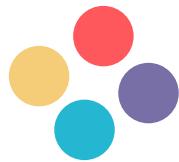
How efficient is DEvIANT?

- Compute an analytic approximate confidence interval rather than an empirical one (Linear to the size of input)

How to prune unpromising sub-search spaces ?

Algorithm

How to efficiently discover the most relevant patterns ?



How efficient is DEvIANT?

- Compute an **analytic approximate confidence interval** rather than an empirical one (Linear to the size of input)

How to prune unpromising sub-search spaces ?

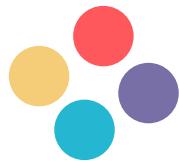
- Tight optimistic estimates by using Eppstein and Hirschberg algorithm. (Linear to the size of input)



Eppstein, D., & Hirschberg, D. S. Choosing subsets with maximum weighted average. Journal of Algorithms, 24(1), 177-193, 1997

Algorithm

How to efficiently discover the most relevant patterns ?



How efficient is DEvIANT?

- Compute an **analytic approximate confidence interval** rather than an empirical one (Linear to the size of input)

How to prune unpromising sub-search spaces ?

- Tight optimistic estimates by using Eppstein and Hirschberg algorithm. (Linear to the size of input)

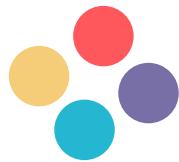


Eppstein, D., & Hirschberg, D. S. Choosing subsets with maximum weighted average. *Journal of Algorithms*, 24(1), 177-193, 1997

- Ensure **nested confidence intervals** property (a condition on min. support threshold) (Linear to the size of input)

Algorithm

How to efficiently discover the most relevant patterns ?



How efficient is DEvIANT?

- Compute an **analytic approximate confidence interval** rather than an empirical one (Linear to the size of input)

How to prune unpromising sub-search spaces ?

- Tight optimistic estimates by using Eppstein and Hirschberg algorithm. (Linear to the size of input)

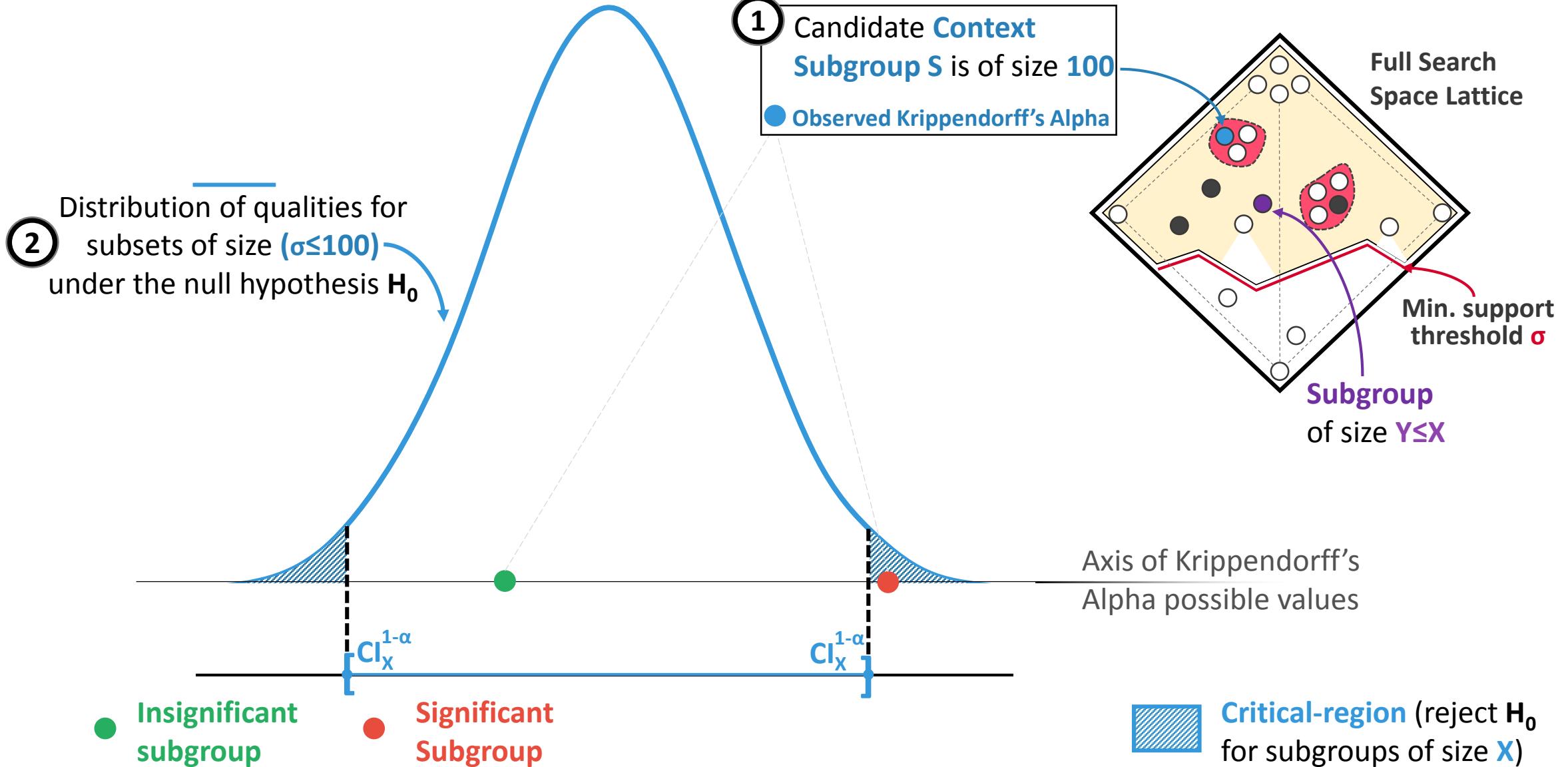


Eppstein, D., & Hirschberg, D. S. Choosing subsets with maximum weighted average. *Journal of Algorithms*, 24(1), 177-193, 1997

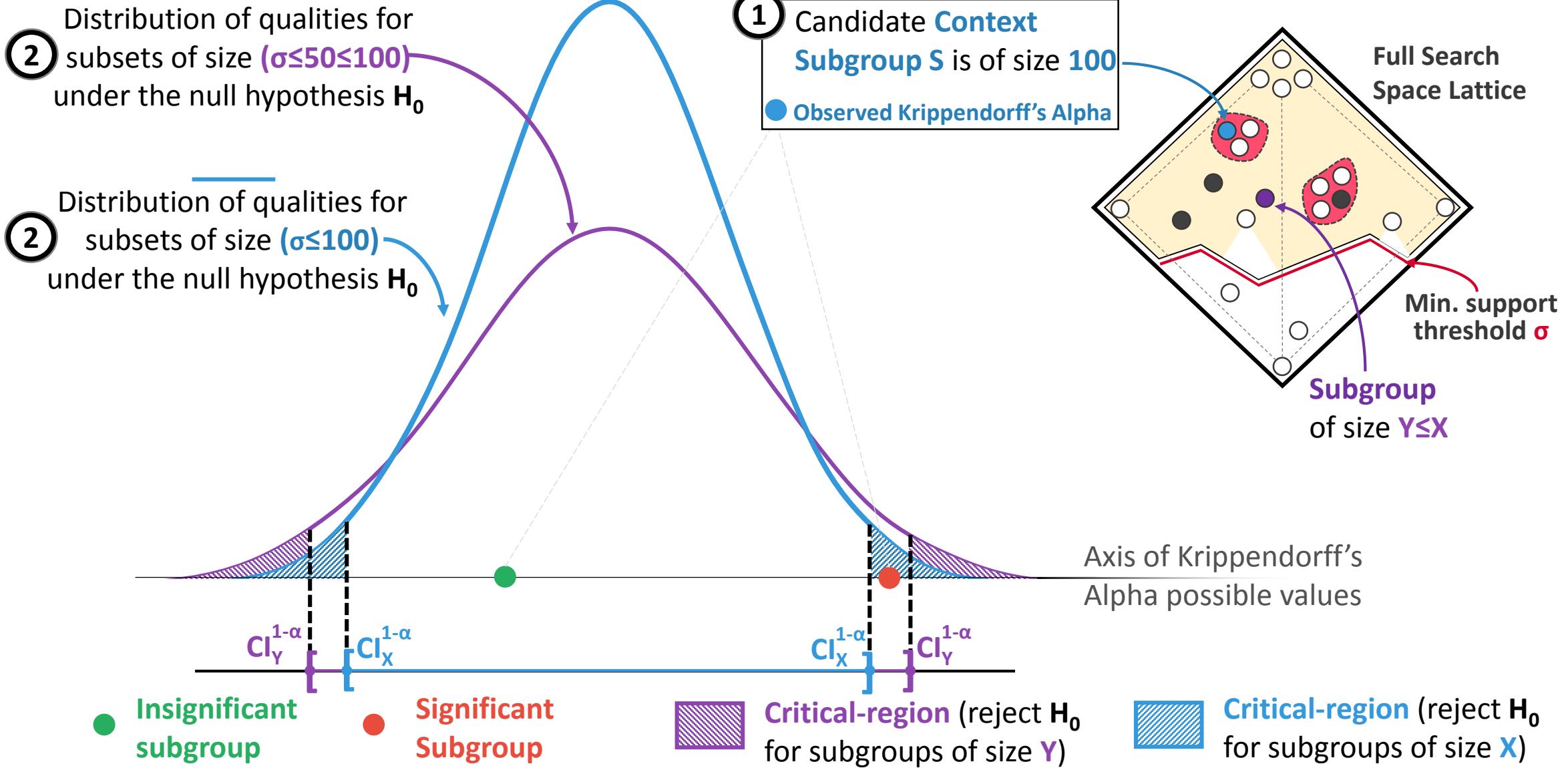
- Ensure **nested confidence intervals** property (a condition on min. support threshold) (Linear to the size of input)

- Prune when the OE region is **inside** the confidence interval.

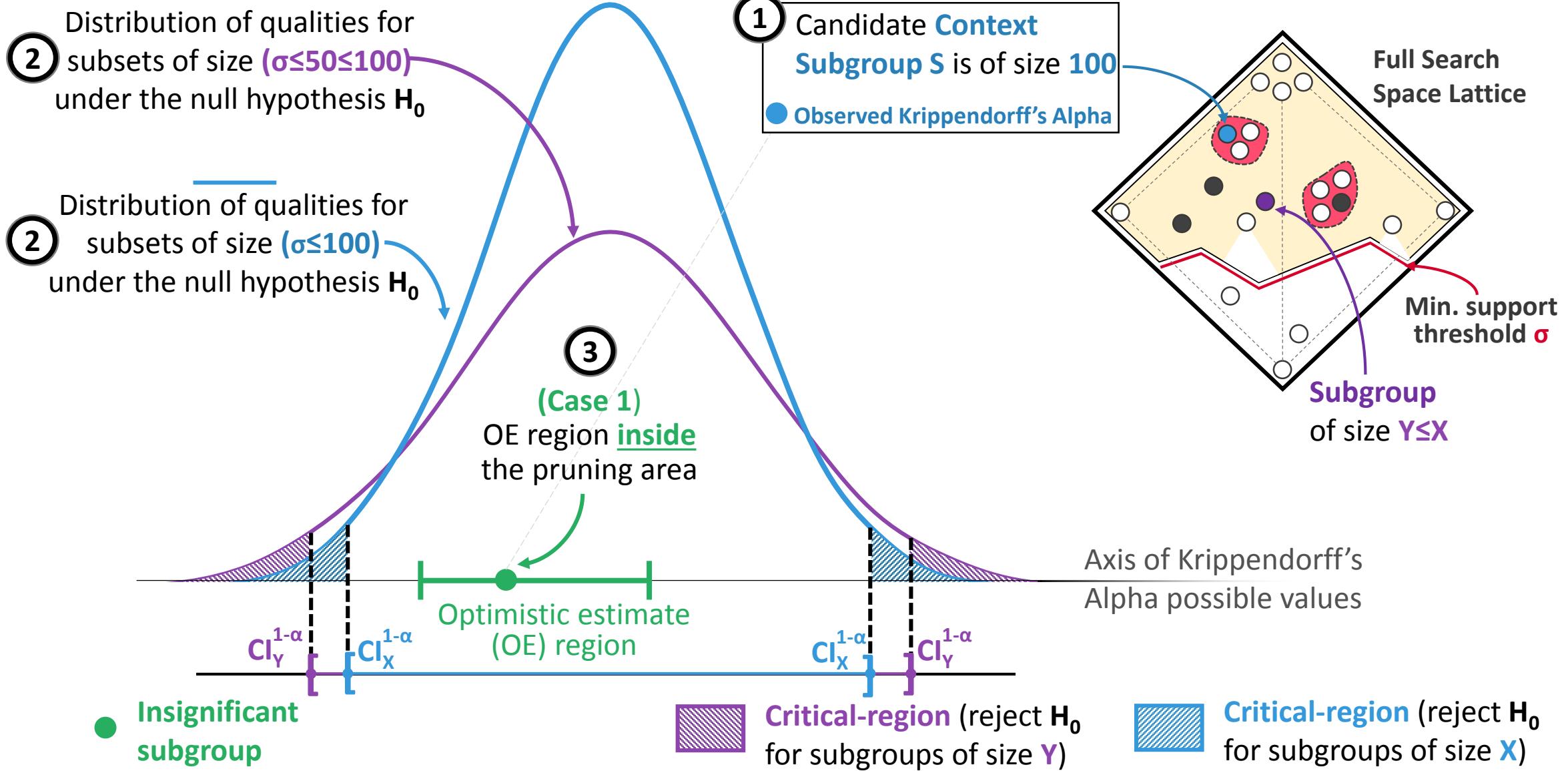
DEvIANT – Pruning Properties (4/4)



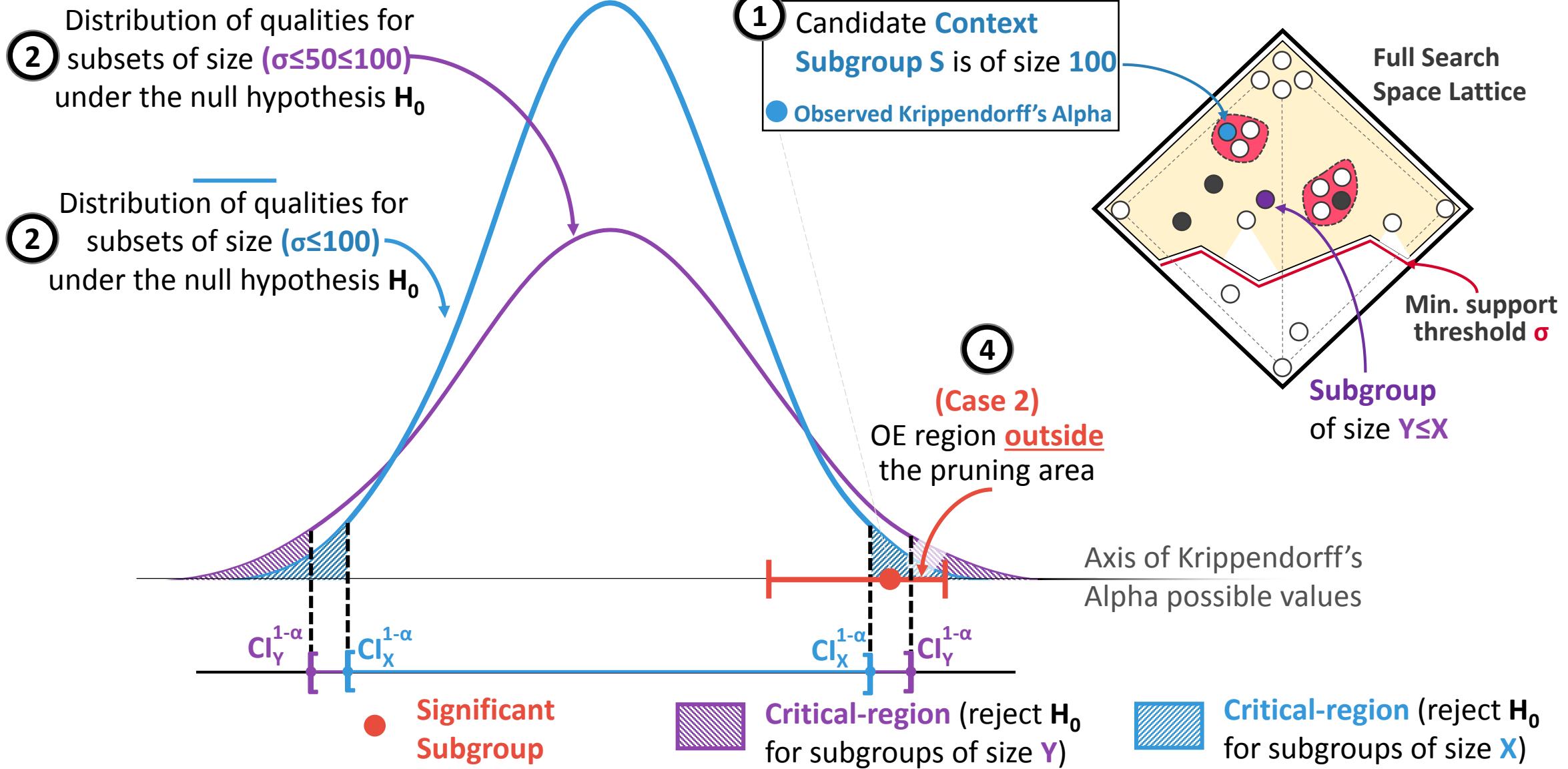
DEvIANT – Pruning Properties (4/4)



DEvIANT – Pruning Properties (4/4)



DEvIANT – Pruning Properties (4/4)





EMPIRICAL EVALUATION

Empirical Evaluation – Some results

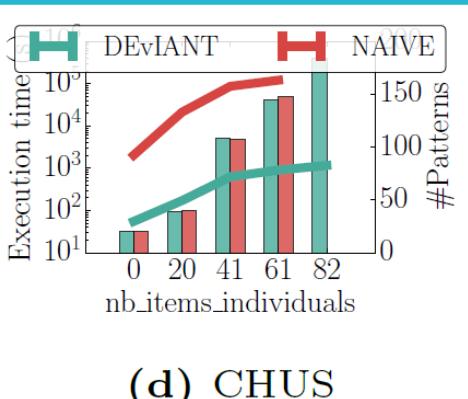
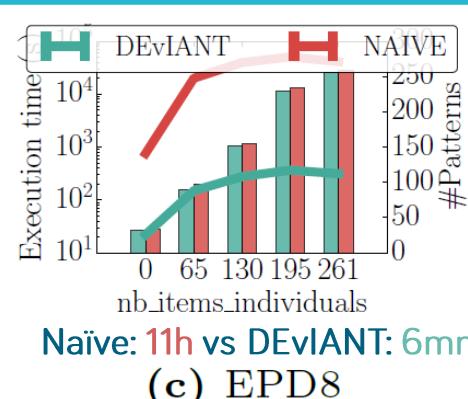
Dataset	Nb entities	Size entities. Attributes	Nb individuals	Size indiv. attributes	Nb outcomes	Sparsity	Outcomes Type
EPD8	4704	437	848	82	3 100 000	78,6%	Categorical
CHUS	17350	307	1373	261	3 000 000	31,2%	Categorical

Dataset	Nb entities	Size entities. Attributes	Nb individuals	Size indiv. attributes	Nb outcomes	Sparsity	Outcomes Type
EPD8	4704	437	848	82	3 100 000	78,6%	Categorical
CHUS	17350	307	1373	261	3 000 000	31,2%	Categorical
Movielens	1681	161	943	27	100 000	6,3%	Numerical
Yelp	127000	851	1 000 000	6	4 150 000	0,003%	Numerical



Bottom line

> x100
Quicker

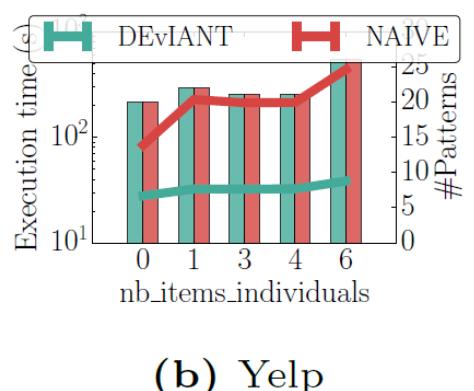
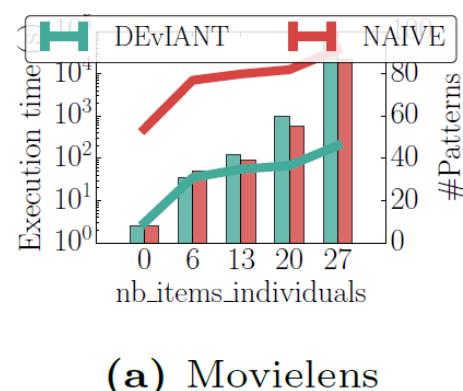
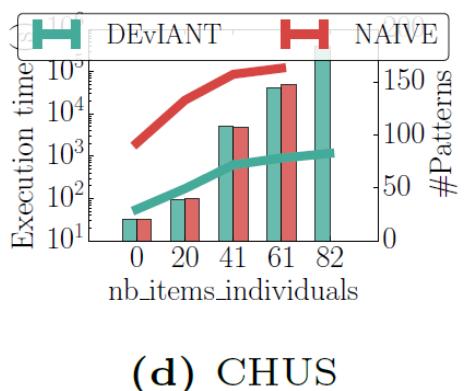
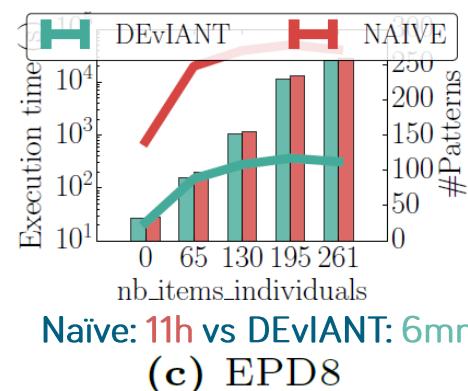


Dataset	Nb entities	Size entities. Attributes	Nb individuals	Size indiv. attributes	Nb outcomes	Sparsity	Outcomes Type
EPD8	4704	437	848	82	3 100 000	78,6%	Categorical
CHUS	17350	307	1373	261	3 000 000	31,2%	Categorical
Movielens	1681	161	943	27	100 000	6,3%	Numerical
Yelp	127000	851	1 000 000	6	4 150 000	0,003%	Numerical



Bottom line

> x 100
Quicker



ANCORE



An online tool to analyze
exceptional behaviors of
parliamentarians

contentcheck.liris.cnrs.fr

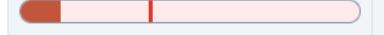
ANCORE



An online tool to analyze
exceptional behaviors of
parliamentarians

contentcheck.liris.cnrs.fr

8th European Parliament (German MEPs)

Agreement-O-Meter	Intra-agreement	Group	Context
 Usual agreement: 0.38 Context agreement: 0.12 P-value: <0.0001	CONFLICTUAL	Germany	[7.10 Free movement and integration of third-country nationals', '7.30 Police, judicial and customs cooperation in general']
 Usual agreement: 0.38 Context agreement: 0.12 P-value: <0.0001	CONFLICTUAL	Germany	[1 European citizenship', '7.10.04 External borders crossing and controls, visas']

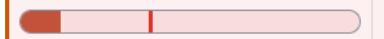
ANCORE



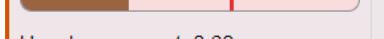
An online tool to analyze
exceptional behaviors of
parliamentarians

contentcheck.liris.cnrs.fr

8th European Parliament (German MEPs)

Agreement-O-Meter	Intra-agreement	Group	Context
 Usual agreement: 0.38 Context agreement: 0.12 P-value: <0.0001	CONFLICTUAL	Germany	['7.10 Free movement and integration of third-country nationals', '7.30 Police, judicial and customs cooperation in general']
 Usual agreement: 0.38 Context agreement: 0.12 P-value: <0.0001	CONFLICTUAL	Germany	['1 European citizenship', '7.10.04 External borders crossing and controls, visas']

8th European Parliament (ECR MEPs)

Agreement-O-Meter	Intra-agreement	Group	Context
 Usual agreement: 0.62 Context agreement: 0.28 P-value: <0.05	CONFLICTUAL	ECR	['3.10.06.01 Fruit, citrus fruits']
 Usual agreement: 0.62 Context agreement: 0.32 P-value: <0.05	CONFLICTUAL	ECR	['2.50.10 Financial supervision', '5.20.03 European Central Bank (ECB), ESCB']

ANCORE



An online tool to analyze **exceptional behaviors** of parliamentarians

contentcheck.liris.cnrs.fr

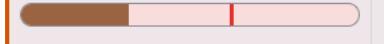
8th European Parliament (German MEPs)

Agreement-O-Meter	Intra-agreement	Group	Context
 Usual agreement: 0.38 Context agreement: 0.12 P-value: <0.0001	CONFLICTUAL	Germany	[7.10 Free movement and integration of third-country nationals', '7.30 Police, judicial and customs cooperation in general']
 Usual agreement: 0.38 Context agreement: 0.12 P-value: <0.0001	CONFLICTUAL	Germany	[1 European citizenship', '7.10.04 External borders crossing and controls, visas']

Movielens (Example)

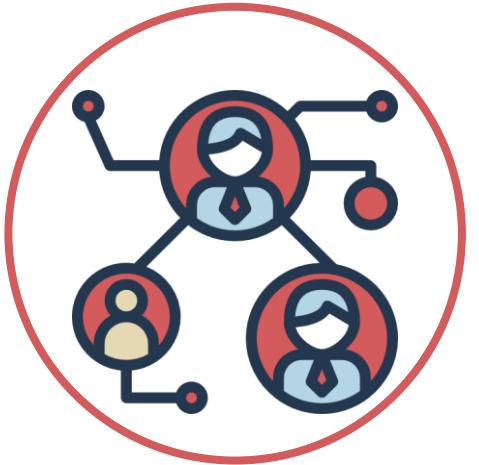
Group	Context	p-value
Old	Action and Adventure and Crime Movies	<0,01 (Conflict)
Middle-Aged Men	Adventure and Musical Movies	<0,01 (Consensus)
Old	Adventure and Musical Movies	<0,01 (Conflict)

8th European Parliament (ECR MEPs)

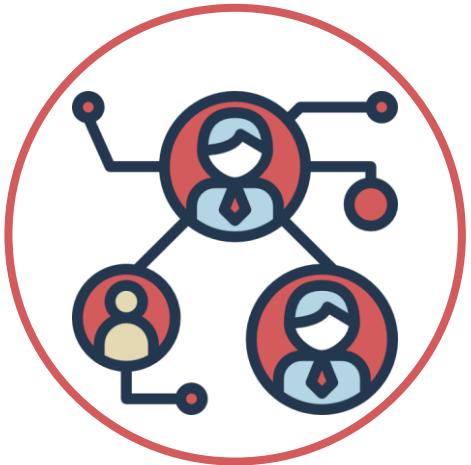
Agreement-O-Meter	Intra-agreement	Group	Context
 Usual agreement: 0.62 Context agreement: 0.28 P-value: <0.05	CONFLICTUAL	ECR	[3.10.06.01 Fruit, citrus fruits']
 Usual agreement: 0.62 Context agreement: 0.32 P-value: <0.05	CONFLICTUAL	ECR	[2.50.10 Financial supervision', '5.20.03 European Central Bank (ECB), ESCB']



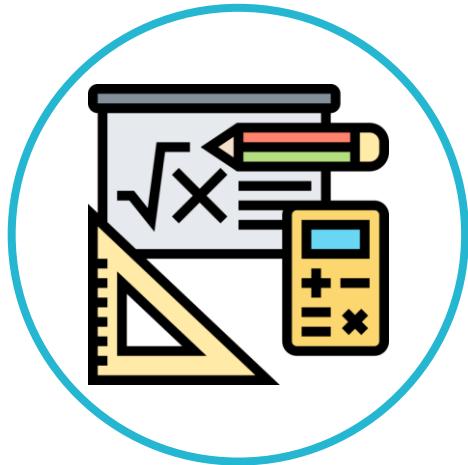
CONCLUSIONS



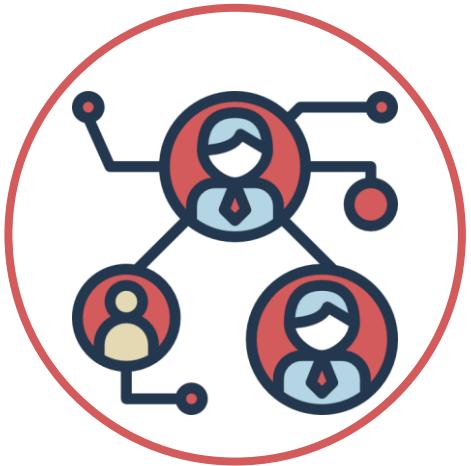
Formalization of the problem
of discovering exceptional
agreements within groups
in Behavioral Data



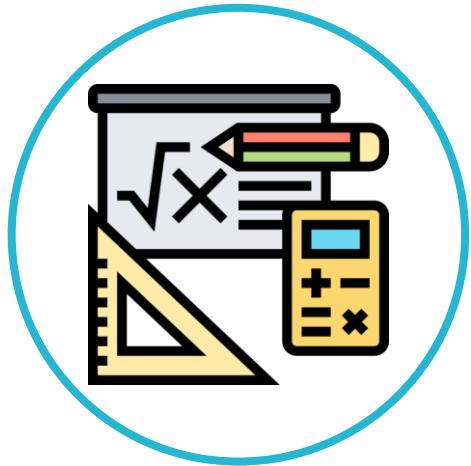
Formalization of the problem
of discovering exceptional
agreements within groups
in Behavioral Data



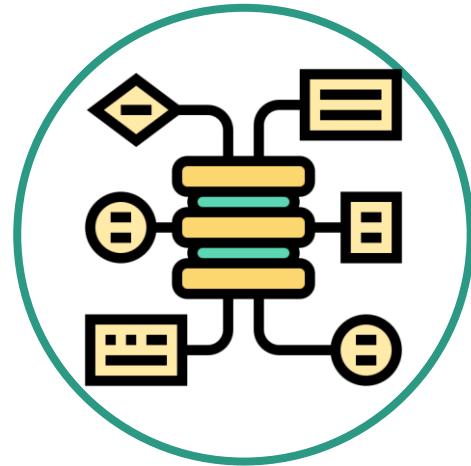
Analytic approximations
for the confidence intervals
for a computationally
efficient significance
assessment



Formalization of the problem
of discovering exceptional
agreements within groups
in Behavioral Data



Analytic approximations
for the confidence intervals
for a computationally
efficient significance
assessment



Safe-pruning
properties for an
efficient branch-and-
bound algorithm

GRAÇIAS

AČIŪ

DĚKUJI

THANKS FOR YOUR TIME

GO RAIBH MAITH AGAT

AČIŪ

TAK

GO RAIBH MAITH AGAT

DZIĘKI

GRACIAS

PALDIES

GRAZIE

TACK

MULŞUMIRI

Ευχαριστώ



QUESTIONS

AITÄH

HVALA

MERCI

KIITOS

OBRIGAD

RCI

KÖSZ

DANKE

HVALA

BEDA

Contact : adnene.belfodil@insa-lyon.fr

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

Looking forward to meeting up with you during the poster session :-)