A local perspective on inconsistency in data-graphs

N. Pardal, V. Martinez

LINDA workshop – 2026

THE UNIVERSITY of EDINBURGH

IIIA-CSIC ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE

CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Talk Roadmap

1. **Introduction**
   - Motivation
   - Graph DBs
   - Consistency
   - Inconsistency measures
   - Navigational logics

2. **Our approach**
   - Two languages
   - Contexts
   - Bases
   - Origin semantics

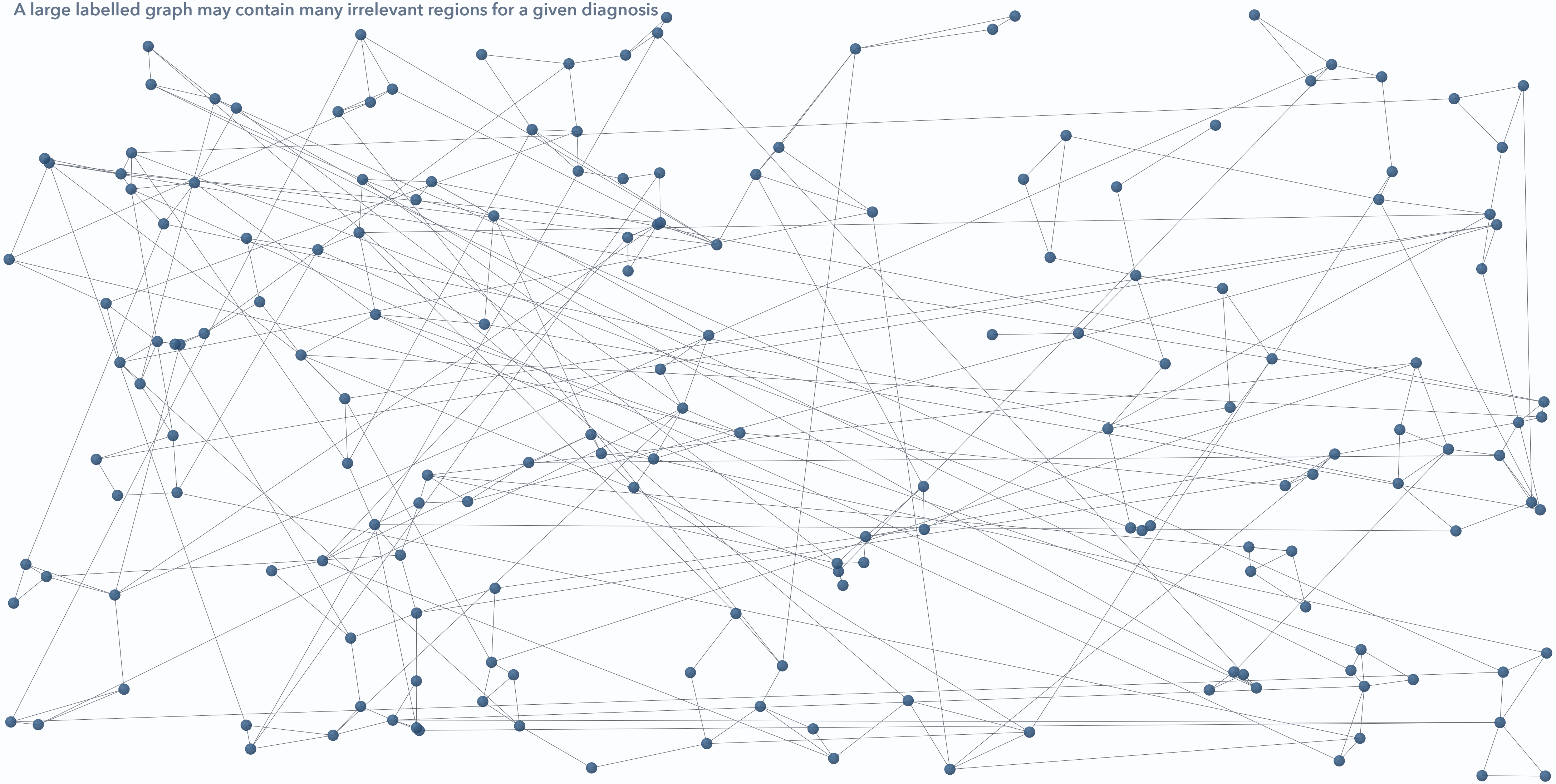
3. **Witnesses and local measures**
   - Problematic pairs/ paths
   - Local witnesses
   - Measuring violations from origins

4. **Postulates**
   - Rationality postulates for local measures

5. **Conclusions and Future Work**

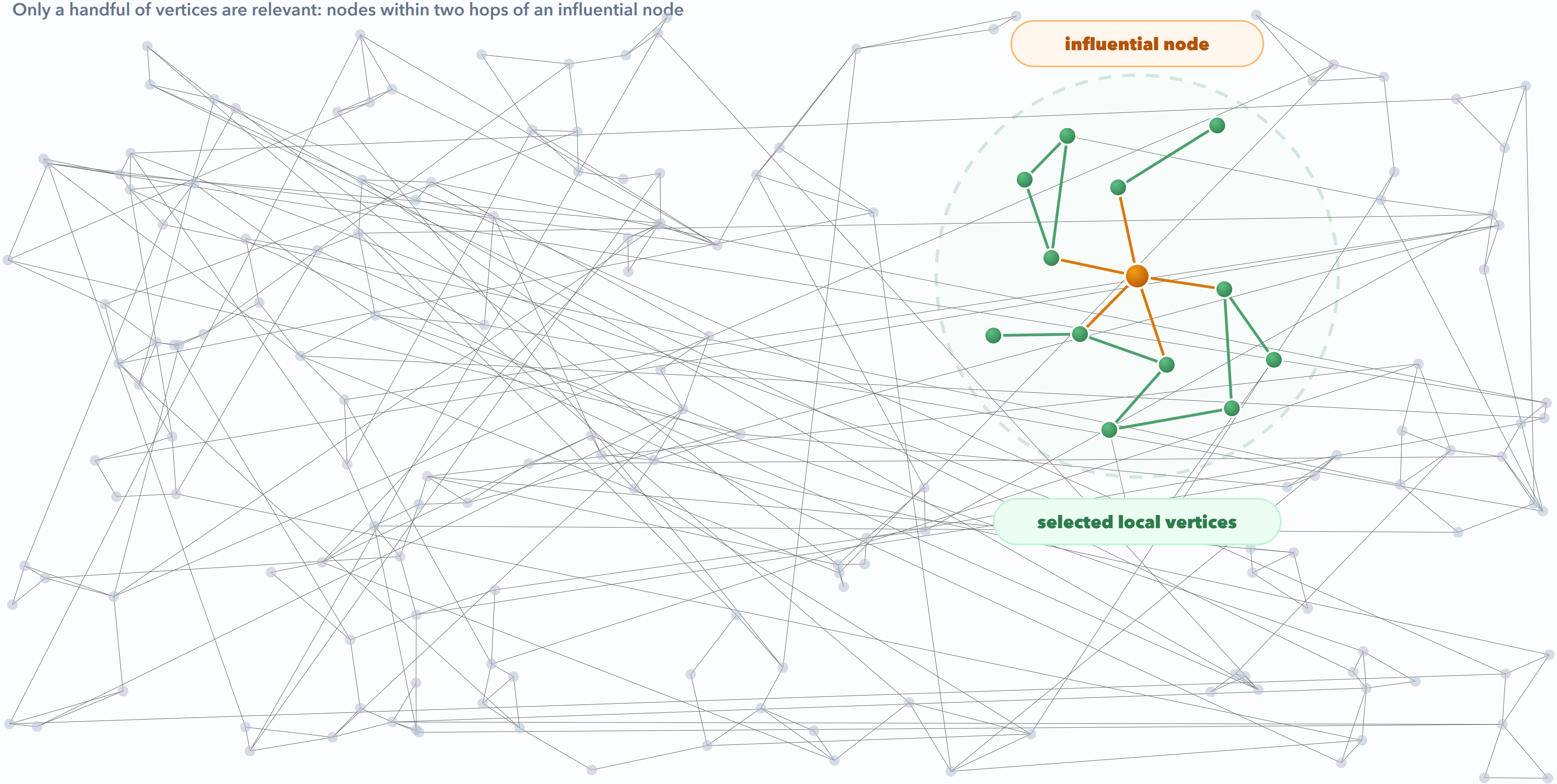
Graph data G

A large labelled graph may contain many irrelevant regions for a given diagnosis



Same graph, local context highlighted

Only a handful of vertices are relevant: nodes within two hops of an influential node



Graph databases

Data as labelled directed graphs

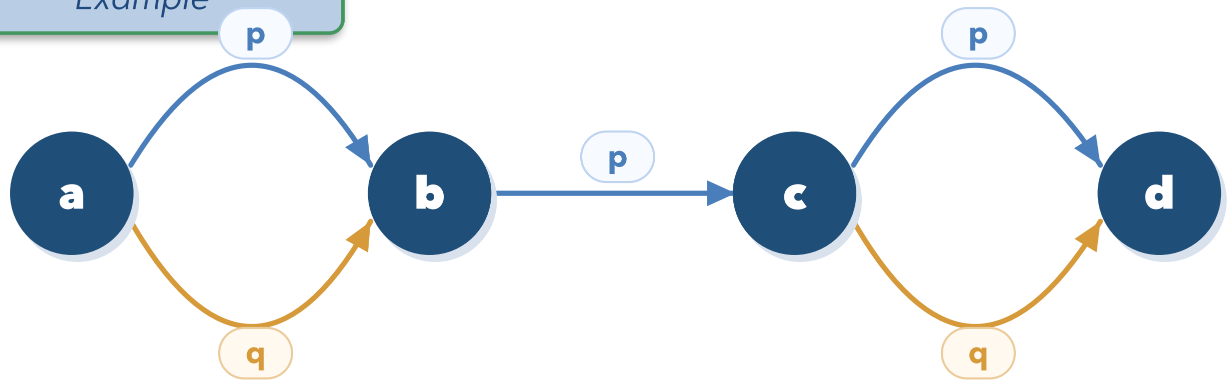
A graph database is a finite directed graph with labels from a given Σ_e on its edges:

$G = (V, E)$, where V = nodes and

$E \subseteq V \times \Sigma_e \times V$ = labelled edges

For us, a graph database *or data-graph* over Σ_n, Σ_e is a tuple $G=(V, E, \{X_p : p \text{ in } \Sigma_n\})$ where $G = (V,E)$ is a traditional edge-labelled graph and X_p is the sets of nodes with label *p*.

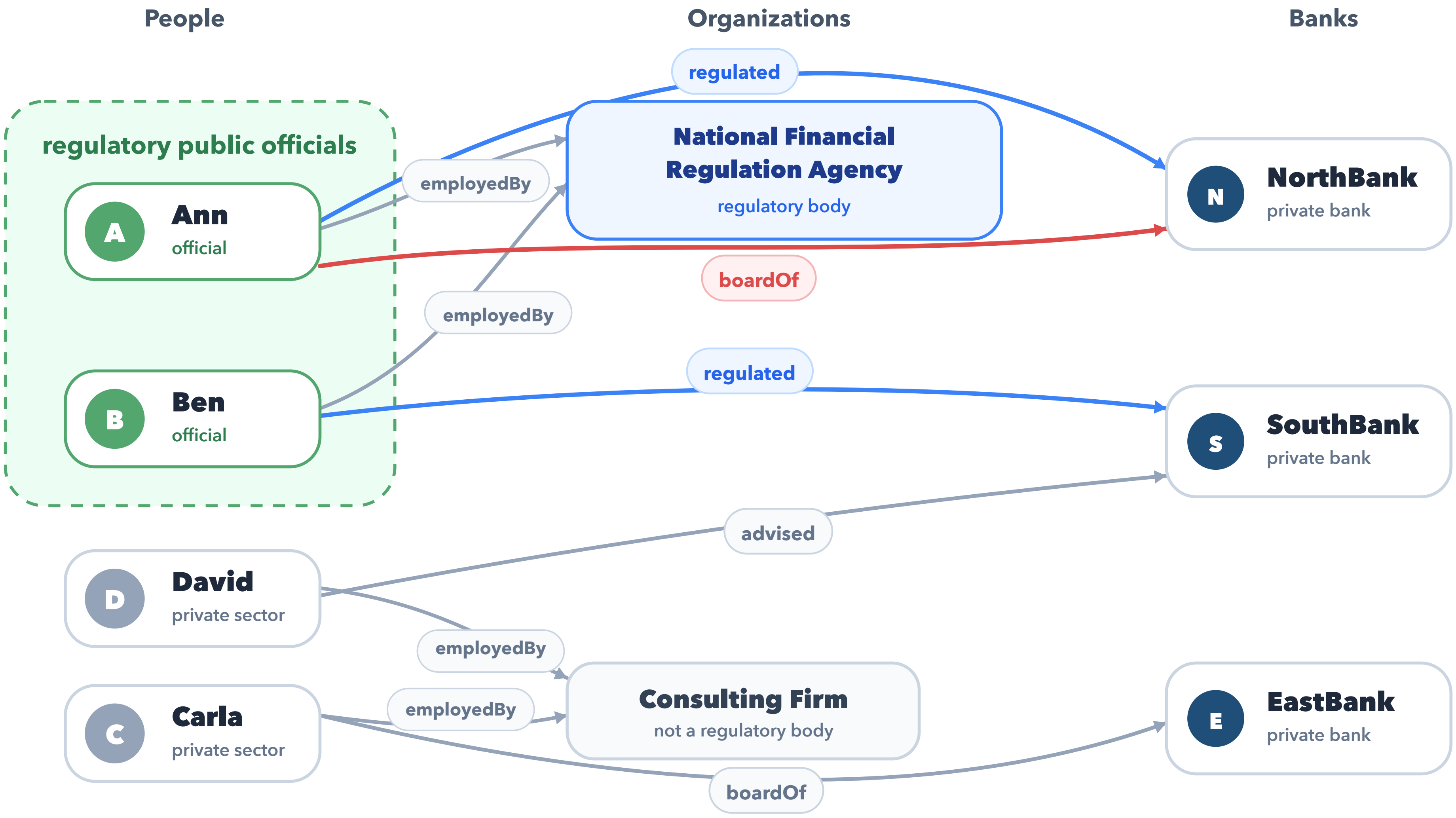
Example



Nodes: a, b, c, d
Edge labels: p, q

Motivating example

Revolving-door conflicts



Conflict-of-interest (inconsistency) rule:
regulated + boardOf

Regular Path Queries

Path queries in the graph

An RPQ is a *regular expression over edge labels.*

It returns pairs of nodes connected by a matching path.

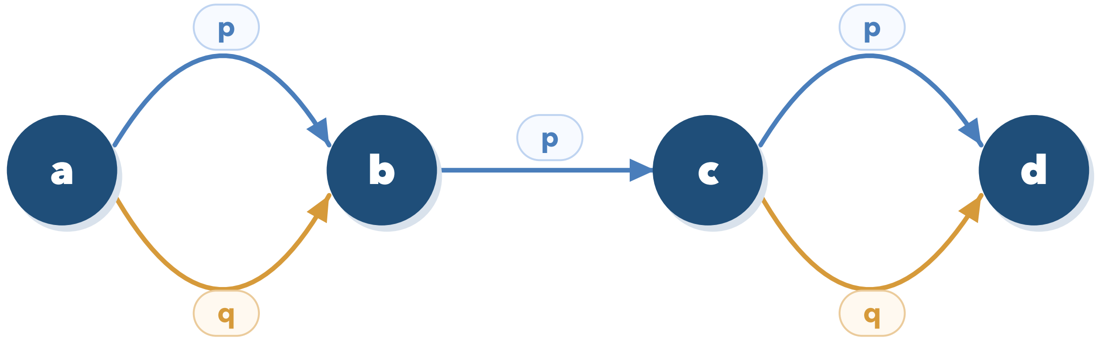
p

q

p · p

p ∪ q

p*



Example query: $p \cdot p$

Matches the pairs (a, c) via **a**—*p*→**b**—*p*→**c**
And (b, d) via **b**—*p*→**c**—*p*→**d**

Integrity Constraints

Integrity constraints are those rules/properties that are desirable for the domain we intend to model.

A graph database is **consistent** if it satisfies all the imposed rules.

Integrity Constraints

Integrity constraints are those rules/properties that are desirable for the domain we intend to model.

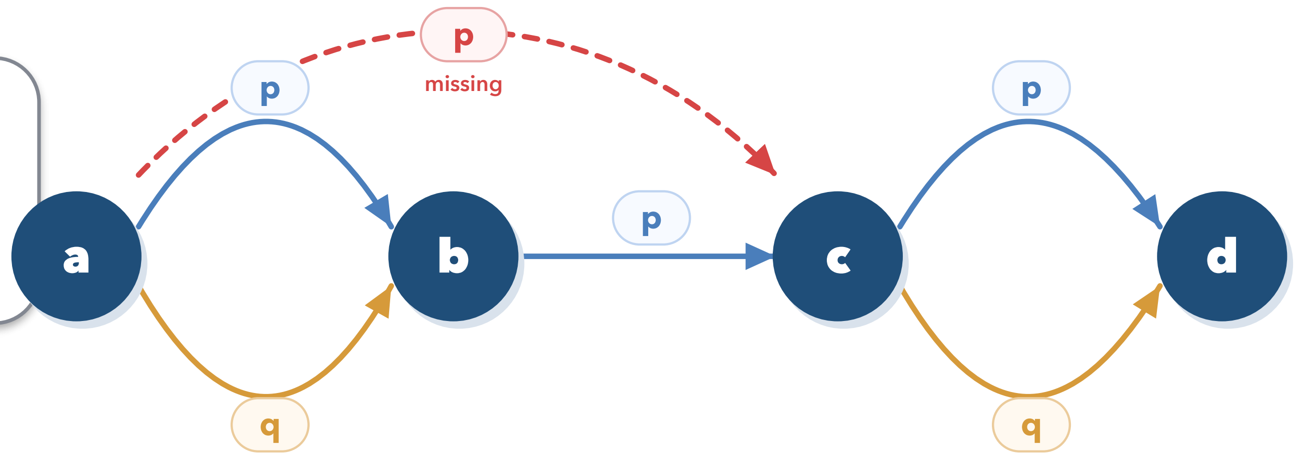
A graph database is **consistent** if it satisfies all the imposed rules.

RPCs are of the form:

$$Q_1 \subseteq Q_2$$

<u>Meaning:</u>
Every Q_1–path should also have a Q_2–path between the same two nodes.

Example: $p \cdot p \subseteq p$



Integrity Constraints

Regular Path Constraints

Integrity constraints are those rules/properties that are desirable for the domain we intend to model.

A graph database is **consistent** if it satisfies all the imposed rules.

RPCs are of the form:

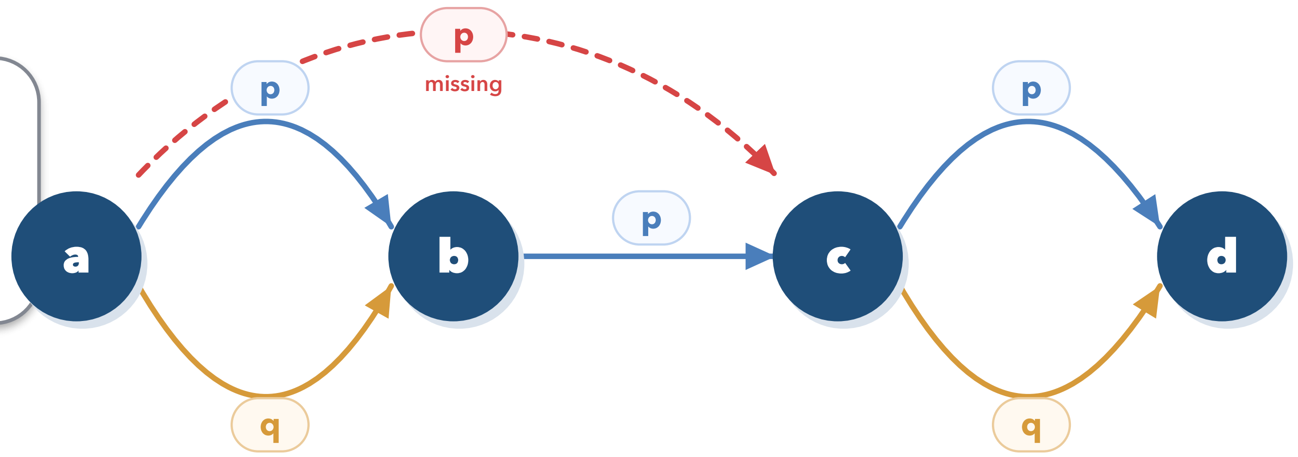
$$Q_1 \subseteq Q_2$$

Meaning:
Every Q_1–path should also have a Q_2–path between the same two nodes.

Example: $p \cdot p \subseteq p$

Problematic pair: (a, c)

There is a p·p path, but no direct p path.



Navigational logics *Path and node constraints*

RPQs

↓

select pairs of nodes

PDL / CPDL

↓

paths + select nodes

Navigational logics

Path and node constraints

RPQs

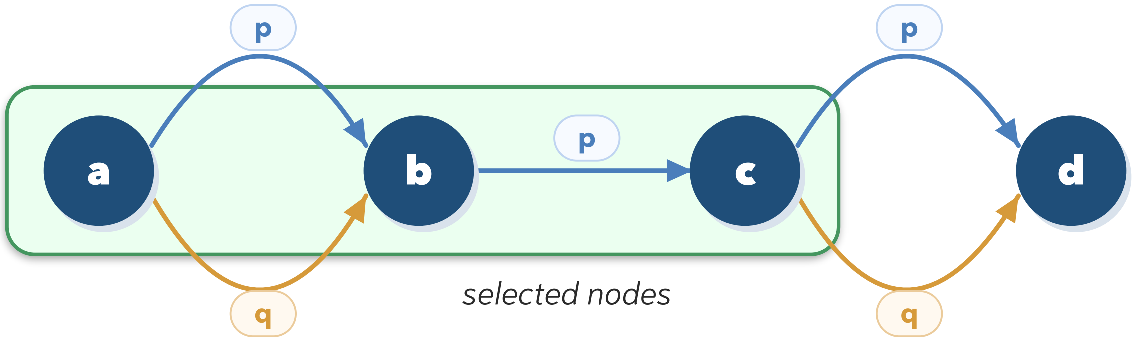
select pairs of nodes

PDL / CPDL

paths + select nodes

Example PDL: $\langle p \cup q \rangle$

nodes with an outgoing p- or q-edge



Selected nodes: {**a**, **b**, **c**}

Inconsistency measures

Global measures for graph databases

An **inconsistency measure** assigns a (non–negative) number to all graphs.

I(G) = 0 iff G is consistent

Examples *(Grant & Parisi, 2024)*

I_B

is there *any* violation?

I_C

how many violated rules?

I_M

MIMS count

Inconsistency measures

Global measures for graph databases

An **inconsistency measure** assigns a (non–negative) number to all graphs.

I(G) = 0 iff G is consistent

Examples

(Grant & Parisi, 2024)

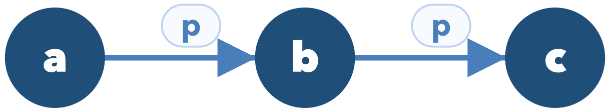
I_B
is there *any* violation?

I_C
how many violated rules?

I_M
MIMS count

MIMS = minimal inconsistent (monotonic) subgraph

$p.p \subseteq q$



Inconsistency measures

Global measures for graph databases

An **inconsistency measure** assigns a (non–negative) number to all graphs.

I(G) = 0 iff G is consistent

Examples

(Grant & Parisi, 2024)

I_B
is there *any* violation?

I_C
how many violated rules?

I_M
MIMS count

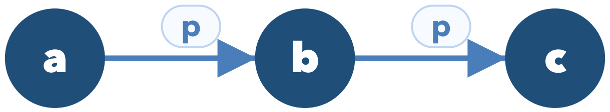
Rationality postulates

Sanity checks for good measures:

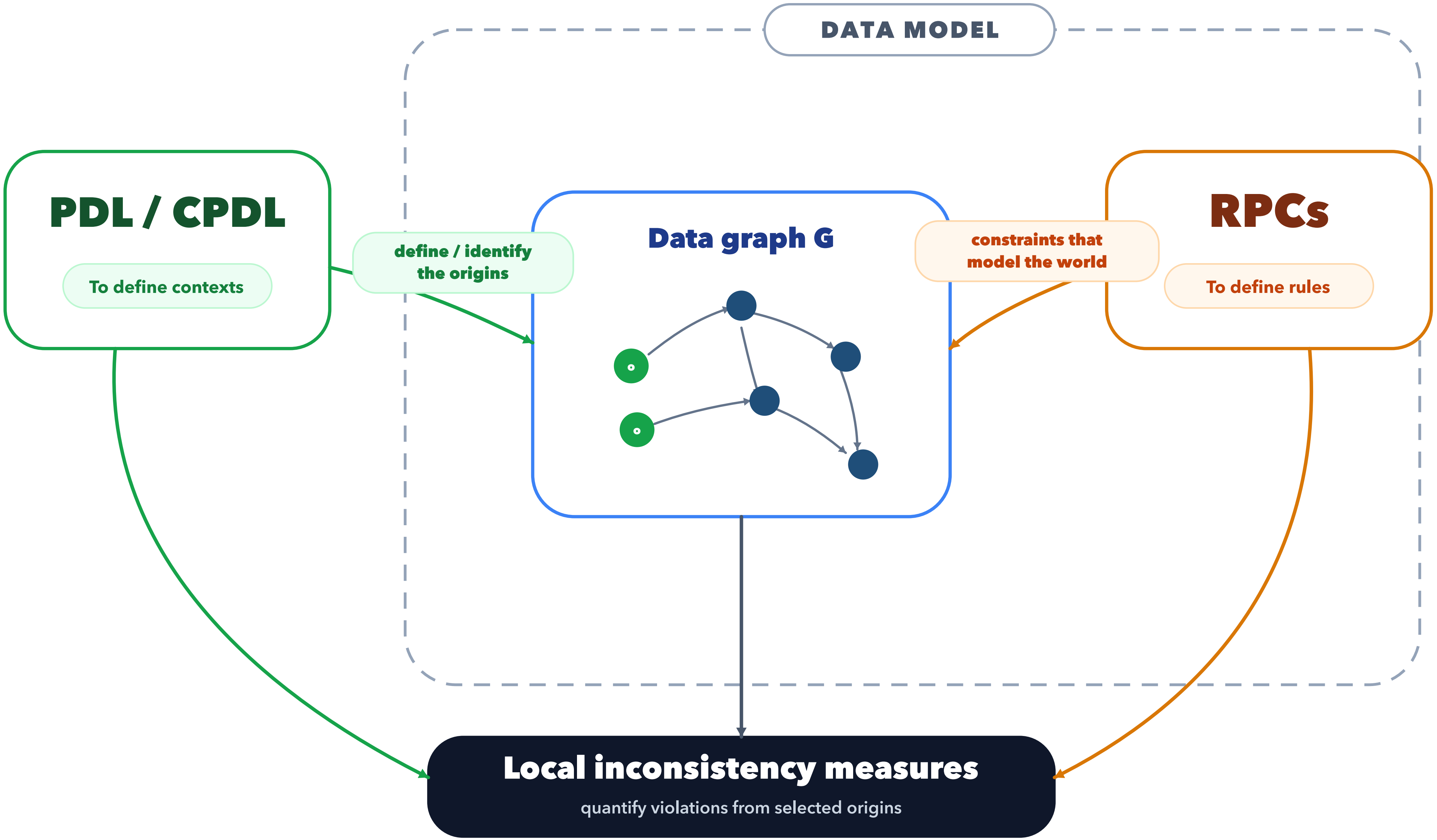
- Monotony
- Independence of free elements
- Penalty for problematic elements
- Super–additivity
- MIMS separability / normalization

MIMS = minimal inconsistent (monotonic) subgraph

$$p.p \subseteq q$$



Our approach


DATA MODEL
PDL / CPDL
To define contexts
define / identify
the origins
Data graph G
constraints that
model the world
RPCs
To define rules
Local inconsistency measures
quantify violations from selected origins

Contexts and Bases

Contextual constraints · Origins

Given some node expressions $\mathcal{C}$ in PDL/CPDL:

Context formulas $\mathcal{C}$

$\mathcal{C}$-context subgraph

maximal subgraph of G whose nodes satisfy all node expressions in $\mathcal{C}$ over G

Base

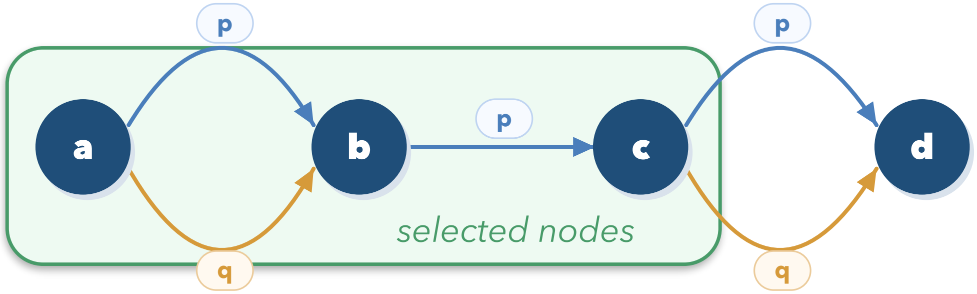
The nodes of the $\mathcal{C}$-context

The context constraints $\mathcal{C}$ define the *origins* $\mathcal{O}$: **nodes of the base**

Example:

PDL / CPDL $\mathcal{C} = \langle p \cup q \rangle$

Question: from which nodes does some p or q edge start?



$\mathcal{C}$ defines the origins as $\mathcal{O}$ = {a, b, c}

$\mathcal{O}$-consistency

Consistency measured from a set of origins

For a finite set of RPCs $\mathcal{R}$ and a set of origins $\mathcal{O} \subseteq V$:

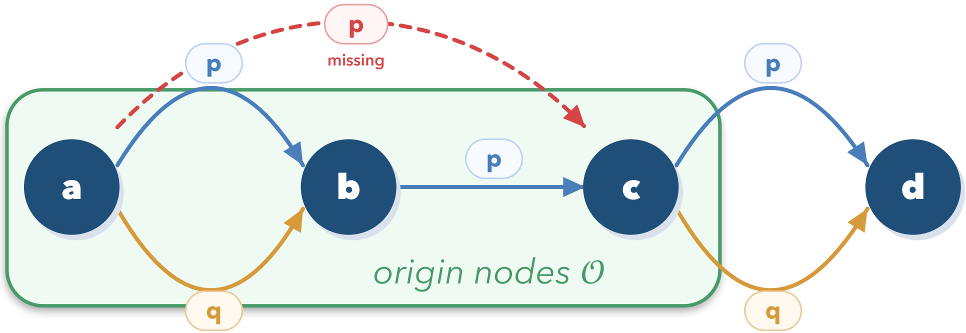
G is $\mathcal{O}$-consistent w.r.t. $\mathcal{R}$ if $\mathcal{O} \times V_G \subseteq [\alpha]_G$ for all $\alpha \in \mathcal{R}$

"If the rules in $\mathcal{R}$ are satisfied by the pairs (o,v) for every o in $\mathcal{O}$"

Example (cont.):

RPC $\mathbf{p \cdot p \subseteq p}$

For every two-step p-path there is a single-step p-path



Does not hold for the origins!

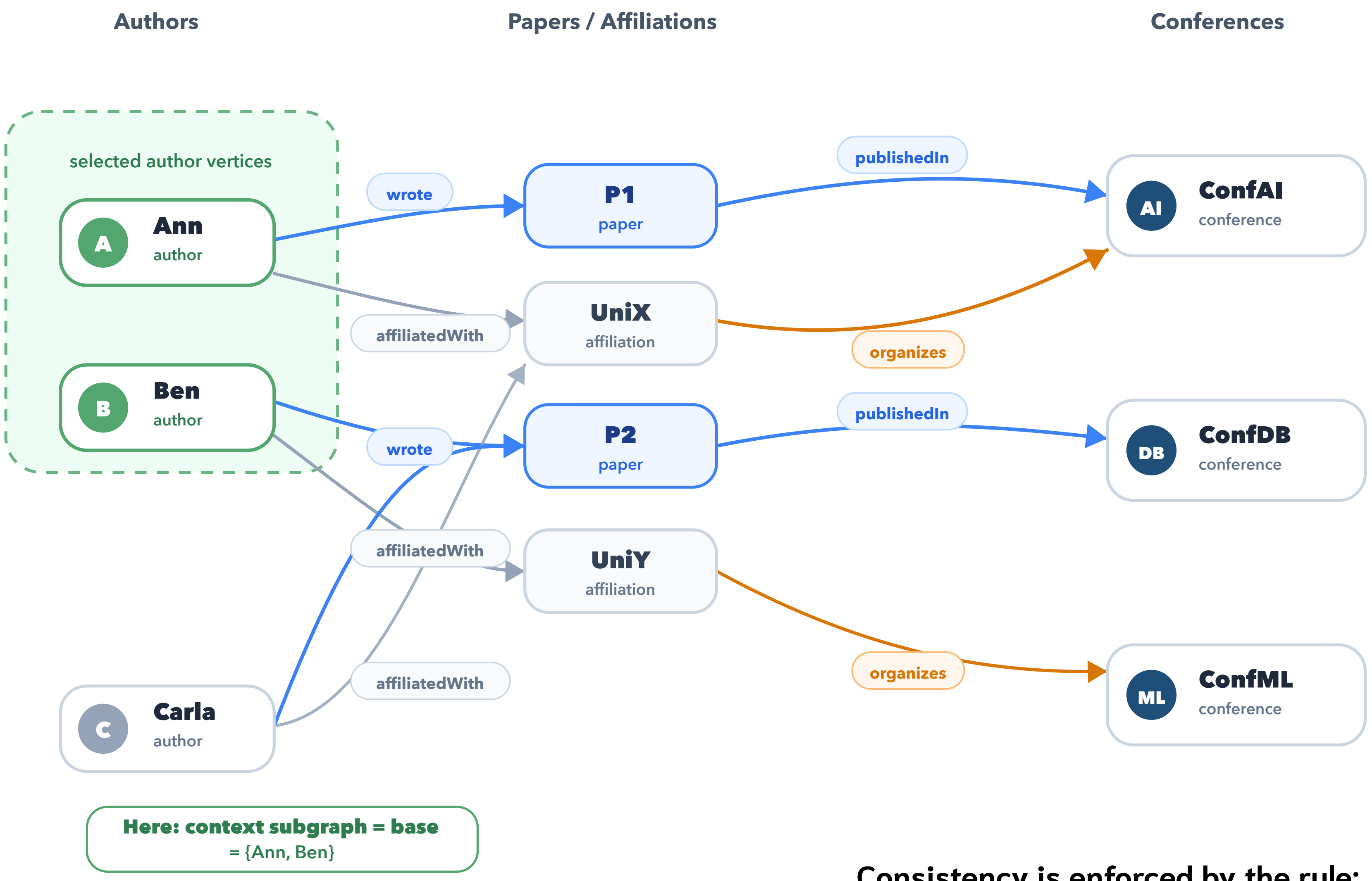
RPQs *define* *<u>what</u>* *to check.* ***PDL/CPDL*** *defines* *<u>where</u>* *to check* *<u>from</u>*.

Used for integrity constraints

Used to generate origins

Another example

Authors, papers, and conferences



Consistency is enforced by the rule:

"a published author cannot have a paper in a conference organized by their affiliation"

Witnesses and local measures

Witnesses and local measures

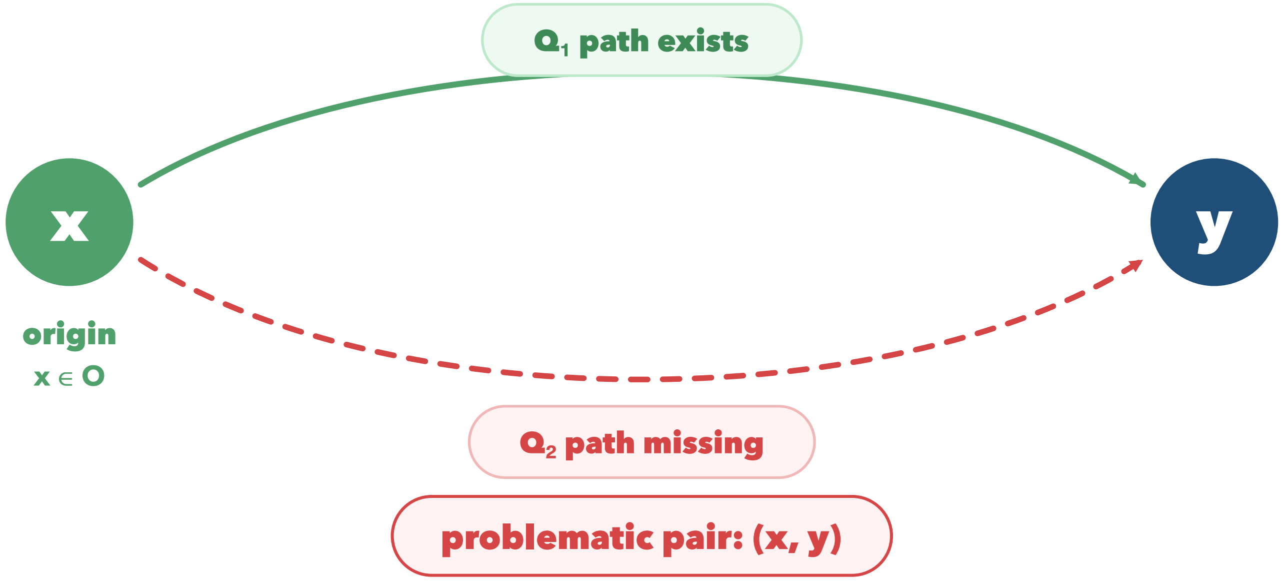
Problematic pairs

Given a rule $r : Q_1 \subseteq Q_2$

A pair (x, y) is problematic when:

1. x is an origin
2. Q_1 connects x to y
3. Q_2 does not connect x to y

Problematic paths are the Q_1-paths that witness such problematic pairs.



Witnesses and local measures

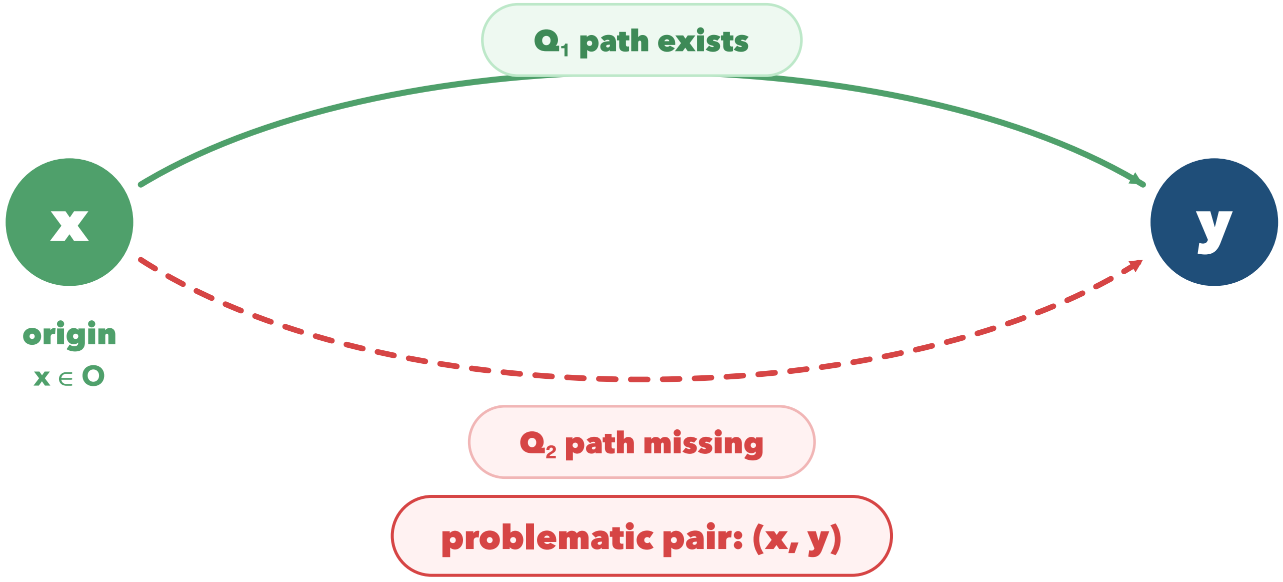
Problematic pairs

Given a rule $r : Q_1 \subseteq Q_2$

A pair (x, y) is problematic when:

1. x is an origin
2. Q_1 connects x to y
3. Q_2 does not connect x to y

Problematic paths are the Q_1-paths that witness such problematic pairs.



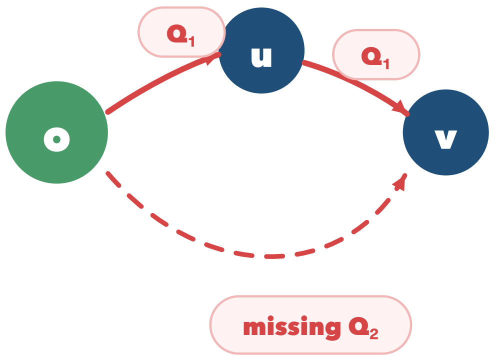
Free elements

Paths, pairs, nodes or edges that do not participate in any local violation.

Minimal witnesses of inconsistency (…or $\mathcal{O}$-MIMS)

Local violation

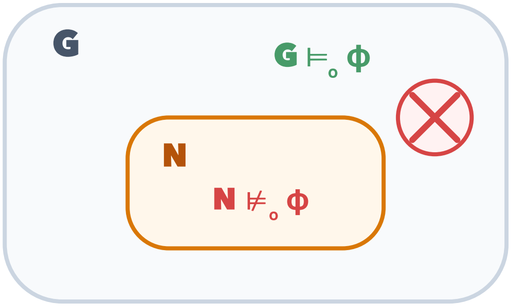
some rule fails from an origin



$N \not\models_o \phi$
for some origin $o \in O$

Monotonicity

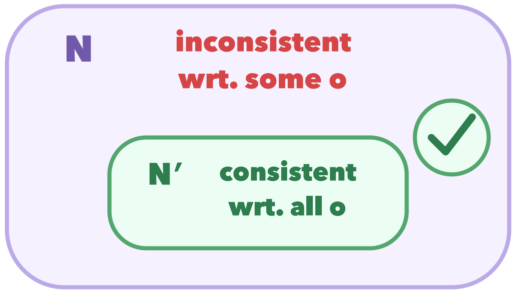
does not create new violations



not allowed:
N violates but G does not

Minimality

no smaller monotonic witness

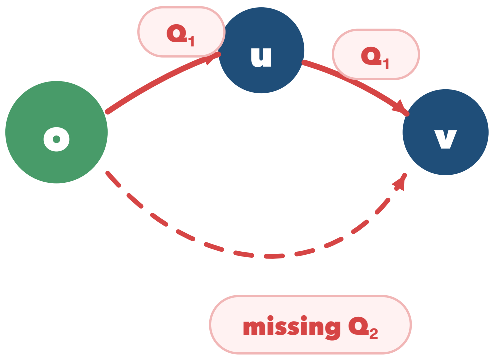


every proper monotonic subgraph
is consistent

Minimal witnesses of inconsistency (…or $\mathcal{O}$-MIMS)

Local violation

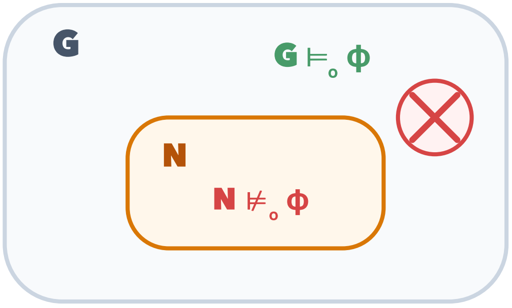
some rule fails from an origin



$N \not\models_o \phi$
for some origin $o \in O$

Monotonicity

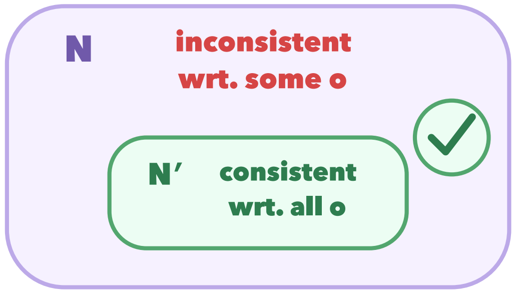
does not create new violations



not allowed:
N violates but G does not

Minimality

no smaller monotonic witness



every proper monotonic subgraph
is consistent

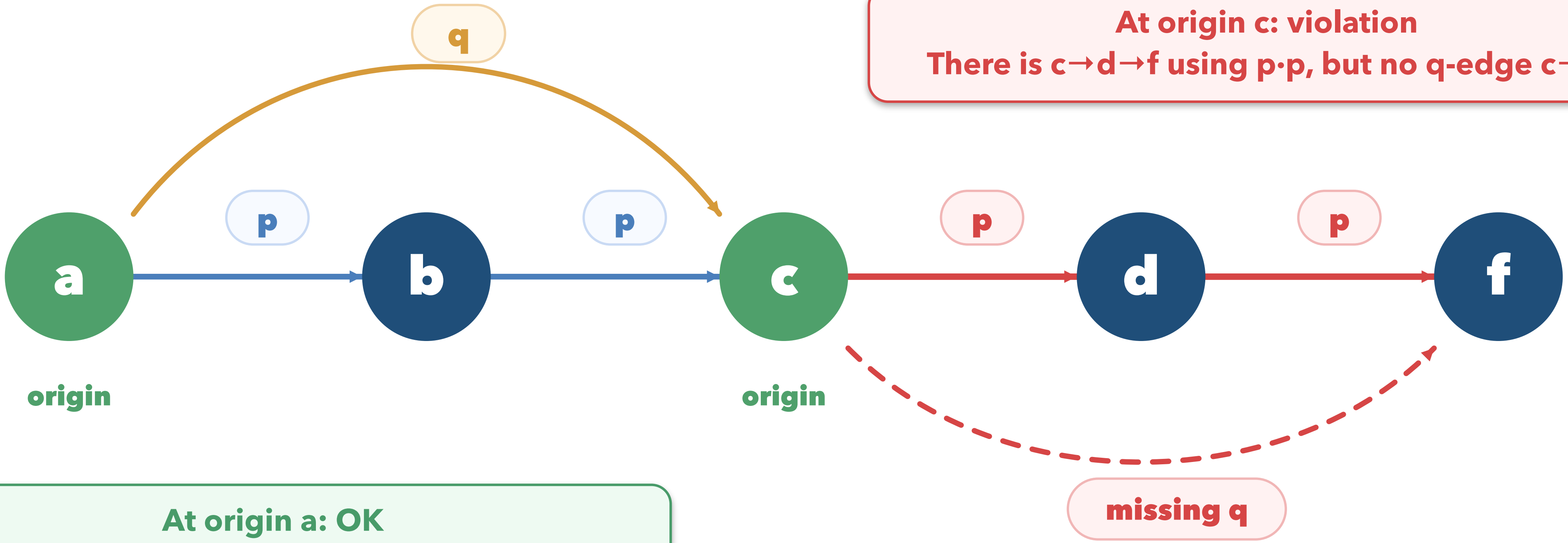
$\mathcal{O}$-MIMS → A minimal monotonic subgraph that contains a violation visible from some origin in $\mathcal{O}$.

Intuition: an $\mathcal{O}$-MIMS is a minimal (local) cause of inconsistency

Example: finding the local witness

Origins $\mathcal{O} = \{a, c\}$; Constraint ϕ: $p \cdot p \subseteq q$

At origin c: violation
There is c→d→f using p·p, but no q-edge c→f.



At origin a: OK
The p·p path a→b→c is matched by a q-edge a→c.

$\mathcal{O}$-MIMS: $M = \{\, c \xrightarrow{p} d,\ d \xrightarrow{p} f \,\}$

Local inconsistency measures

Setup: $\mathcal{O} = V(W_C)$ · R = integrity constraints

These measures are all constraint-based!

***All local measures satisfy:* $I(G,\mathcal{O}) = 0$ *iff* G is $\mathcal{O}$–consistent**

Measure	Definition	What does it ask?
I_B	1 iff G is not $\mathcal{O}$–consistent	Is there at least one violation starting from the selected origins?
$I_{B'}$	$\lvert\mathcal{O}'\rvert/\lvert\mathcal{O}\rvert$	What fraction of the origins witness at least one violation?
I_C	$\max_o \lvert C_o\rvert$	What is the maximum number of rules violated by a single origin?
$I_{C'}$	$\lvert\bigcap_{\{o\,:\,C_o \neq \varnothing\}} C_o\rvert$	(At least) How many rules are violated by every problematic origin?
$I_{C''}$	$\lvert\bigcup_o C_o\rvert$	How many distinct rules are violated from at least one origin?

$\mathcal{O}'$ = { origins that violate at least one rule } · C_o = { rules violated from origin o }

Witness-based measures

Using $\mathcal{O}$-MIMS to count minimal local causes

Recall that $\mathcal{O}$-MIMS...

o → violating pattern → MIMS

origin

1. Are monotonic: do *not* create new violations that are not in G.
2. Violate *some* rule in R, when evaluated from *some* origin o in $\mathcal{O}$.
3. Are minimal: they contain no *inconsistent monotonic* subgraphs

Measure	Definition	What does it ask?
I_M	# $\mathcal{O}$-MIMS in G	How many minimal witnesses of local inconsistency exist for the origin set $\mathcal{O}$?
$I_{M'}$	$\max_o$ |o-MIMS(G)|	What is the largest number of minimal witnesses seen from a single origin?

Rationality Postulates

Rationality Postulates

Classical principles, relative to origin sets

Origin-aware version: every comparison is made with respect to selected origins $\mathcal{O}$

Postulate	Intuition	Origin-aware condition
Monotony	Adding information should not reduce inconsistency if local violations survive.	$G' \subseteq_m^{\mathcal{O}} G \Rightarrow I(G',\mathcal{O}) \leq I(G,\mathcal{O})$
Edge Independency	Removing a free edge should not change the local measure.	e free w.r.t. $\mathcal{O} \Rightarrow I(G',\mathcal{O}) = I(G,\mathcal{O})$
Vertex Independency	Removing a free vertex should not change the local measure.	v free w.r.t. $\mathcal{O} \Rightarrow I(G',\mathcal{O}) = I(G,\mathcal{O})$
Edge Super-Additivity	Edge-disjoint (monotonic) sources of inconsistency should accumulate.	$G_i \subseteq^{\mathcal{O}i}{}_m G_1 \cup G_2$ and $E_1 \cap E_2 = \emptyset \Rightarrow$ $I(G_1 \cup G_2, \mathcal{O}_1 \cup \mathcal{O}_2) \geq I(G_1, \mathcal{O}_1) + I(G_2,\mathcal{O}_2)$
Vertex Super-Additivity	Vertex-disjoint graphs should accumulate their inconsistencies.	$V_1 \cap V_2 = \emptyset \Rightarrow$ $I(G_1 \cup G_2, \mathcal{O}_1 \cup \mathcal{O}_2) \geq I(G_1, \mathcal{O}_1) + I(G_2,\mathcal{O}_2)$

$\mathcal{O}$-MIMS postulates

Rationality principles for minimal local witnesses

$\mathcal{O}$-MIMS Separability

Modularity

If the witnesses of the union are exactly the witnesses from the parts, then the measure is additive.

$\mathcal{O}$-MIMS Super-Additivity

Witnesses survive in the union

If witnesses from each part remain witnesses in the union, then the union has at least the sum.

$\mathcal{O}$-MIMS Normalization

One big witness

If the only minimal witness is the whole graph, the measure assigns value 1.

Separability gives us modularity, but might be too much to ask:

$\mathcal{O}_1$-$MIMS(G_1)$

$\mathcal{O}_2$-$MIMS(G_2)$

→

$\mathcal{O}_1 \cup \mathcal{O}_2$-$MIMS(G_1 \cup G_2)$

separability:
exactly the same witnesses

super-additivity:
same witnesses + maybe more

With origins, the union may reveal cross-witnesses that were not visible in either part alone.

Conclusions and Future Work

Conclusions and Future Work

- A local perspective to measure consistency in graphs, based on origins, decoupling node relevance and ICs.
- Preliminary analysis of how to adequate classical/graph inconsistency measure notions to our local notions.
- New measures, preliminary analysis of postulates and properties.

Conclusions and Future Work

- A local perspective to measure consistency in graphs, based on origins, decoupling node relevance and ICs.
- Preliminary analysis of how to adequate classical/graph inconsistency measure notions to our local notions.
- New measures, preliminary analysis of postulates and properties.

- Analyze differences with the global setting
- Find measures/properties that reflect and better exploit the locality
- Study computational complexity
- Empirical testing of these measures