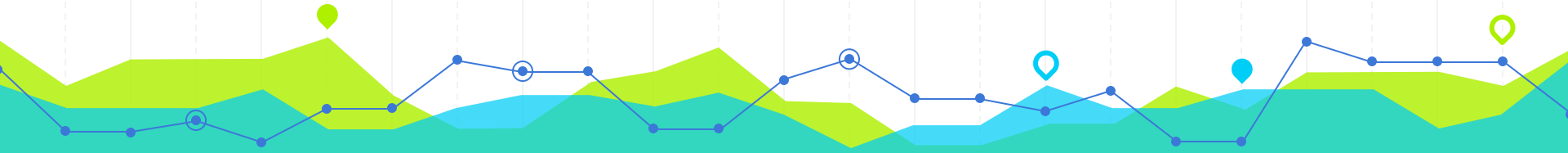
cirad LIRMM SIFR project Agence Nationale de la Recherche ANR

Subjective and generic semantic distances in ViewpointS

I am Guillaume Surroca

surroca@lirmm.fr
With my co-authors :
Philippe Lemoisson, Clément Jonquet and Stefano A. Cerri



Introduction

1

Let’s talk about knowledge representations and semantic distances

INTRODUCTION

Two ways of elicitating knowledge

CONCENSUS

- **Rich** semantics created by circle of experts
- **Slow** evolution rate due to the **costly elicitation** of Semantic Web knowledge structures

EMERGENCE

- **Poor** semantics emerging from interactions
- **Fast** evolution rate due to the profusion of interactions on the Social Web

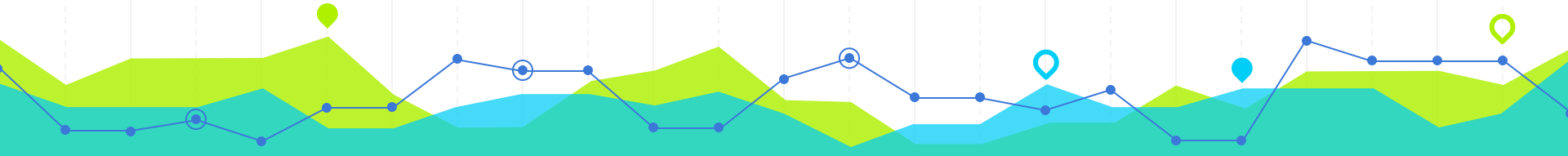
Our Objective : To draw the benefits of both approaches by creating a knowledge representation formalism capable of integrating

ACCORDING TO OUR OBJECTIVE WE NEED A FORMALISM …

- … catching the subjectivity of Social Web knowledge
- … giving value to the structure that characterize Semantic Web datasets

The integration of Social Web and Semantic Web may allow for a new synergy, lowering the cost of data and raising the computational value of gathering it.

T. Gruber in [Collective knowledge systems: Where the Social Web meets the Semantic Web]



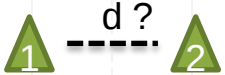
IF WE DIVIDE WEB RESOURCES IN THREE CATEGORIES

Knowledge Supports
Ex: documents, databases, videos

Knowledge Providers
Ex: web users, software agents (data mining)

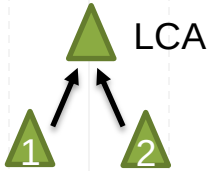
Knowledge Descriptors
Ex: ontology concepts, folksonomy tags

THE LITTERATURE* PROPOSES SEMANTIC RELATEDNESS MEASURES



Beetween knowledge descriptors

Based on Information Content
Ex: Resnik

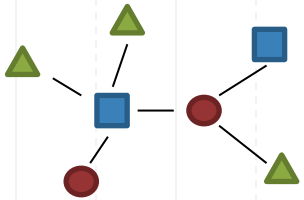


Based on Taxonomical structure
Ex: Rada, Leacock & Chodorov

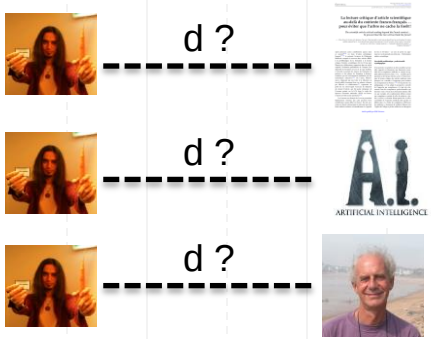
* According to our current knowledge

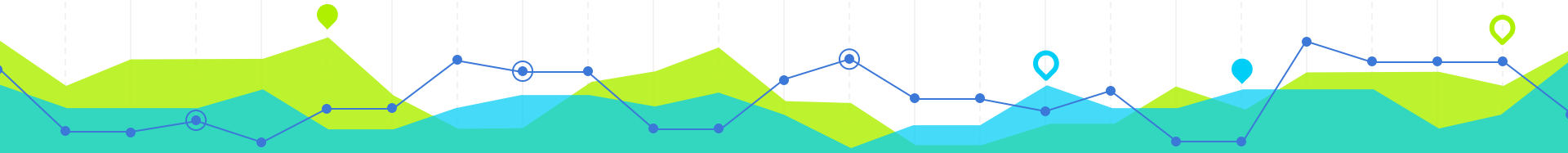
WE NEED A SEMANTIC MEASURE

◉… which not depends on taxonomical structure



◉… able to work genericly on knowledge providers, descriptors and supports



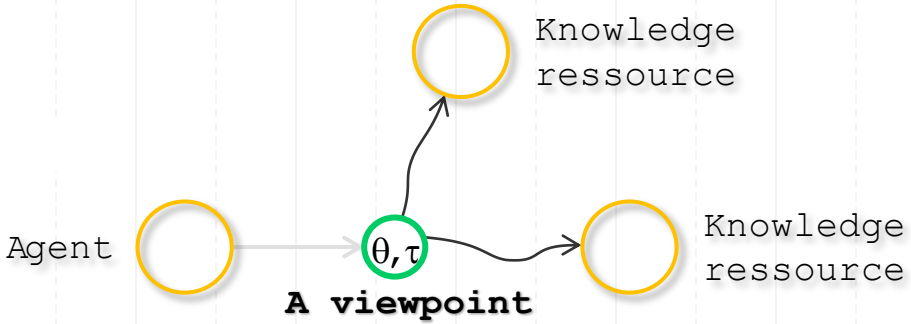


The ViewpointS approach

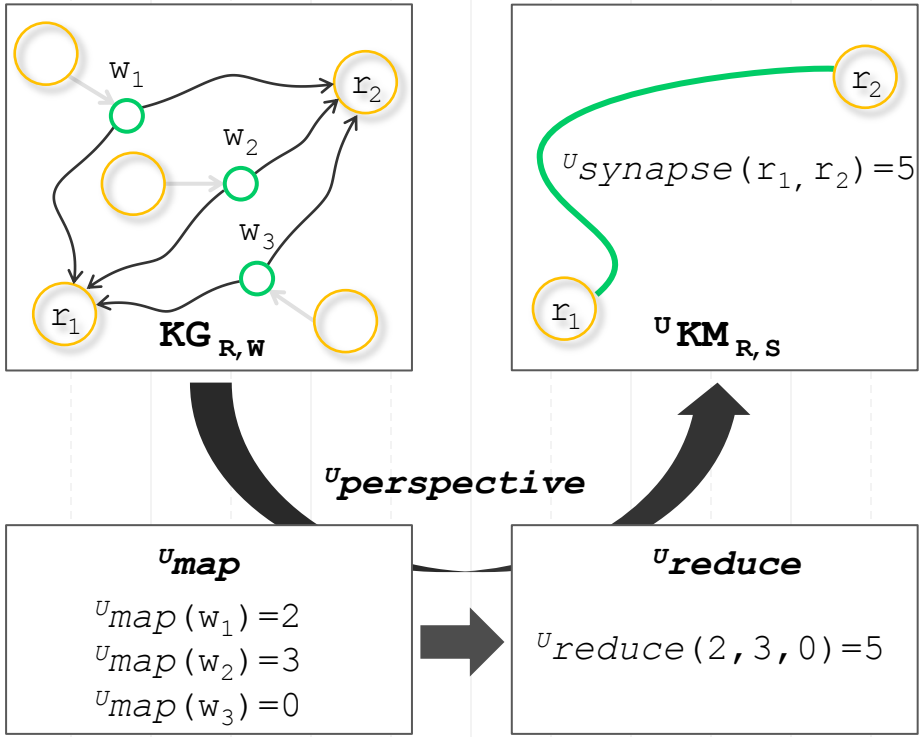
2

The knowledge representation formalism and
the semantic distance measures we apply on it

FIRST DEGREE OF SUBJECTIVITY



SECOND DEGREE OF SUBJECTIVITY



SEMANTIC DISTANCES ON VIEWPOINTS

Shortest Path Distance

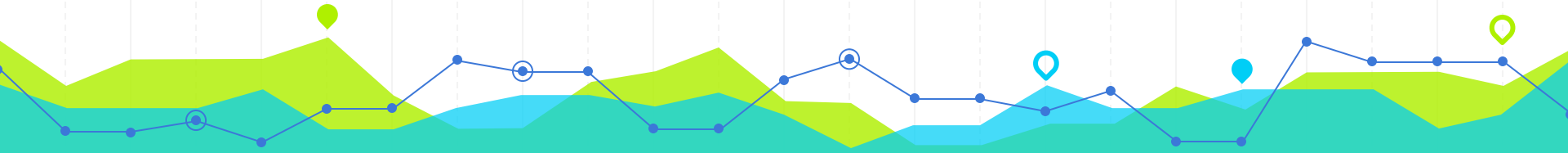
- **Fast** semantic distance method based on the lengh of the shortest path
- **Ignores multiple paths** beetween two resources
- Considering n synapses with respective strengths s_1 to s_n

$$d_{SPN}(r_1, r_2) = \sum_{i=1}^{n} \frac{1}{s_i}$$

Multiple Path Distance

- **Slow** semantic distance method
- **Considers all the bounded paths** beetween two resources
- Considering n paths beetween r_1 and r_2 with respectives lenghts l_1 to l_n

$$d_{MPN}(r_1, r_2) = \frac{1}{\frac{1}{l_1} + \frac{1}{l_2} + \frac{1}{l_3} + \cdots + \frac{1}{l_n}}$$



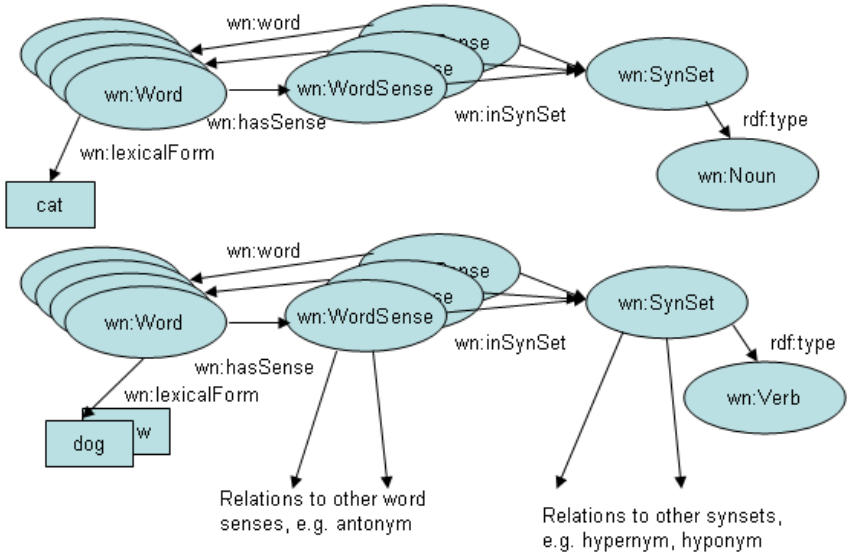
Evaluation

3

Let’s compare semantic distances in ViewpointS to existing ones

METHOD : Dataset

WordNet 3.0



METHOD : Gold standard

WordSim 353

- 353 words compared by humans

Word 1	Word 2	Human (mean)
love	sex	6.77
tiger	cat	7.35
tiger	tiger	10.00
book	paper	7.46
computer	keyboard	7.62
computer	internet	7.58
plane	car	5.77
train	car	6.31
telephone	communication	7.50

METHOD : Compared methods

Lin

$$sim_{Lin}(c_1, c_2) = \frac{2 \times IC_{Resnik}(MICA(c_1, c_2))}{IC_{Resnik}(c_1) + IC_{Resnik}(c_1)}$$

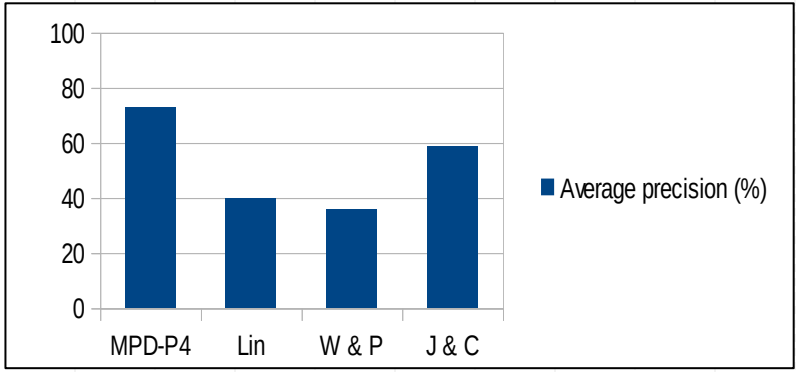
*MICA : Most Informative Common Ancestor
[Dekang Lin. 1998. An Information-Theoretic Definition of Similarity. (July 1998), 296–304]

Wu & palmer

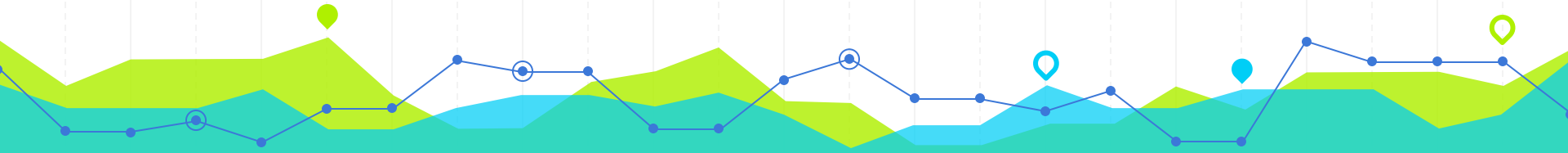
$$sim_{WuPalmer}(c_1, c_2) = \frac{2 \times depth(LCA)}{depth}$$

[Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In Proceedings of the 32nd annual meeting on Association for Computational Linguistics -. Morristown, NJ, USA: Association for Computational Linguistics, 133–138.]

RESULTS



Benchmark summary using wordsim 353 gold standard



Conclusion 4

CONCLUSION

- We presented ViewpointS: a knowledge representation formalism able to **integrate Social Web interactions and Semantic Web data**
- We presented semantic distance measures on ViewpointS: **generic methods which can be specialized** to specific usecases due to the Perspective mechanism
- We presented the Perspective mechanism : all methods are **subjective methods** tied to a specific viewpoints interpretation policy

THANKS!

Any questions?

You can find me at
surroca@lirmm.fr