MULTI-SITE DIVERSIFIED SEARCH AND RECOMMENDATION

LABEX

**Maximilien Servajean
(PhD. Student)**
INRIA & LIRMM
Université de Montpellier
servajean@lirmm.fr

**Esther Pacitti
(Thesis Advisor)**
INRIA & LIRMM
Université de Montpellier
pacitti@lirmm.fr

**Sihem Amer Yahia
(Collaborator)**
CNRS, LIG
sihem.amer-yahia@imag.fr

**Pascal Neveu
(Co-advisor)**
INRA & Supagro
pn@supagro.fr

**NUMEV
Solutions
Numériques
Matérielles et
Modélisation pour
l'Environnement et
le Vivant**

POSITIONNEMENT NUMEV : AXE DONNÉES

KEYWORDS : recommendation, top-k, profile diversity, distributed data management, plant phenotyping, citizen science

ABSTRACT: Many scientific fields produce and consume a considerable amount of diverse data (*e.g.* biology , astronomy, physics) stored in different heterogeneous sites. We investigate two different use cases:

1. In the domain of plant phenotyping, there has recently been increasing interests in finding diverse data coming from different research communities.

2. In botany, the emergence of citizen sciences has fostered the creation of large and structured communities of nature observers. In this context, there is a need to retrieve diverse plant observations from a diverse spectrum of plant families, genus and species.

We investigate profile diversity, a novel idea in searching and recommending scientific items (*e.g.* documents, datasets).

MODEL APPROACH: *PROFDIV*

ProfDiv returns relevant and diversified items shared by relevant and diversified users:

$$score_{DivRSci}(it, u, q) = rel(it, q) \times div_c(it|\{it_1, ..., it_{i-1}\}) \times div_p(u_{it}|\{u_{it_1}, ..., u_{it_{i-1}}\})$$

ProfDiv uses a new Profile Diversification score:

$$div_p(u_{it}|\{u_{it_1}, ..., u_{it_{i-1}}\}) = \frac{1}{N} \times \sum_{v_n \in u_{it_i}} \left[rel_{trust}(v, u, q) \times \prod_{v_m \in \{u_{it_1}, ..., u_{it_{i-1}}\}} (1 - red_p(v_m|v_n)) \right]$$

Profile Diversification enables to return items shared by trustworthy and diversified users with respect to the query's initiator and to the query itself

USE CASE 1: PLANT PHENOTYPING

q=*plant model*		
Undiversified Profiles		
Documents	Communities	Disciplines
Doc_1: short-term resp...	Ecophysiologist	Biologist
Doc_2: Drought and Absc...	Ecophysiologist	Biologist
Doc_3: Control of leaf...	Ecophysiologist	Biologist
Doc_4: The importance of...	Ecophysiologist	Biologist
Diversified Profiles		
Doc_1: short-term resp...	Ecophysiologist	Biologist
Doc_5: A Multiscale Mod...	Modeling	Computer scientists
Doc_2: Drought and Absc...	Ecophysiologist	Biologist
Doc_6: Computational an...	Modeling	Computer scientists

Table 1: Search and recommendation example with profile diversity

Profile diversity enables to retrieve documents shared by users from different communities and disciplines

USE CASE 2: CITIZEN SCIENCE - BOTANY

q = *asteraceae*		
Un-diversified	Profile Diversity	
Élodie Dujardin	Marie Dupont	Pierre Durand

Table 1: Search and recommendation example with profile diversity.

Un-diversified results only contain very redundant plant observations. Profile Diversity enables to return a broader spectrum of diversified plants observations, useful to understand the plant biodiversity

PlantRT: MULTI-SITE SEARCH AND RECOMMENDATION
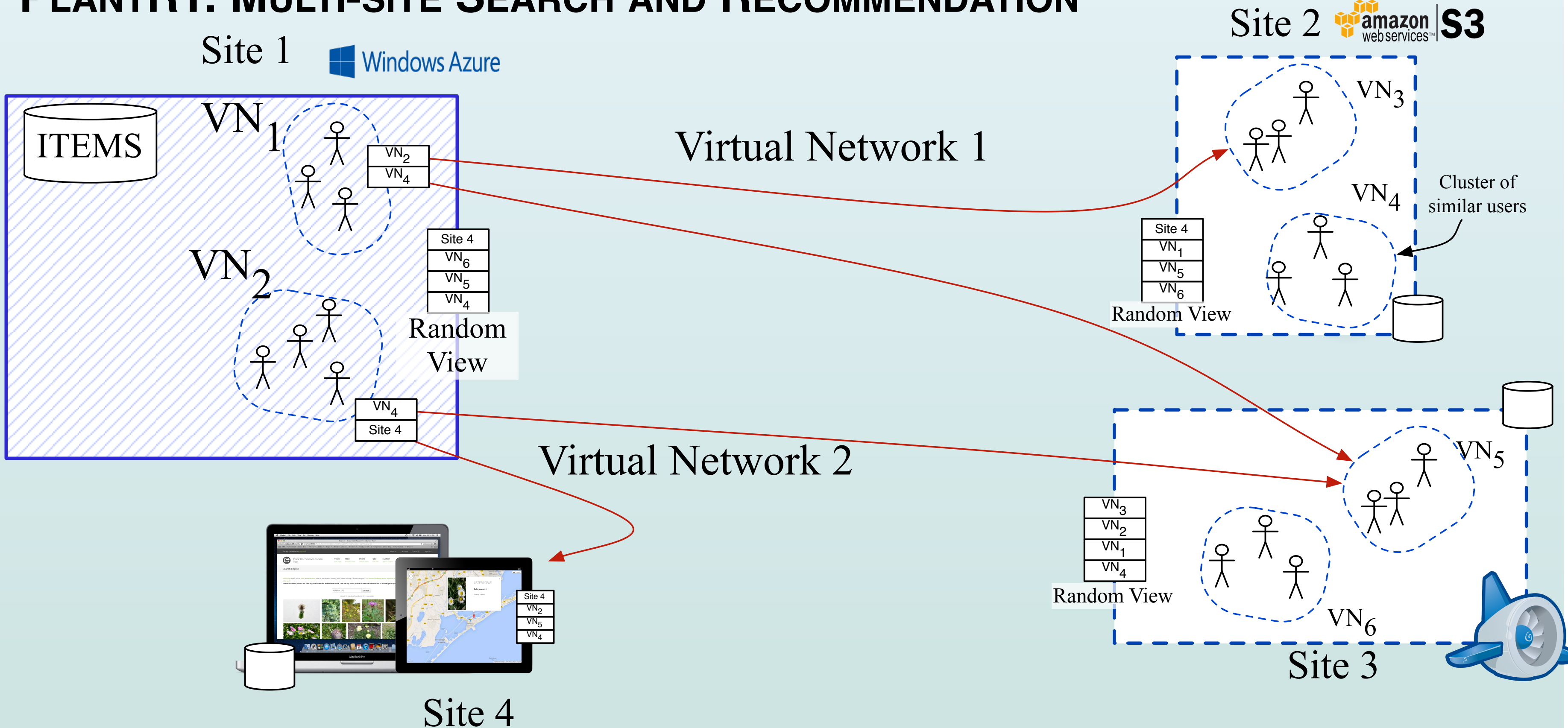
Virtual nodes (VN) groups similar user profiles and are connected through a diversified virtual network. Whenever a user submits a query *q*, it is forwarded to the most diverse virtual nodes. Our prototype is available at the following address:
http://www2.lirmm.fr/~servajean/prototypes/plant-sharing/plant-rt.html

RESULTS
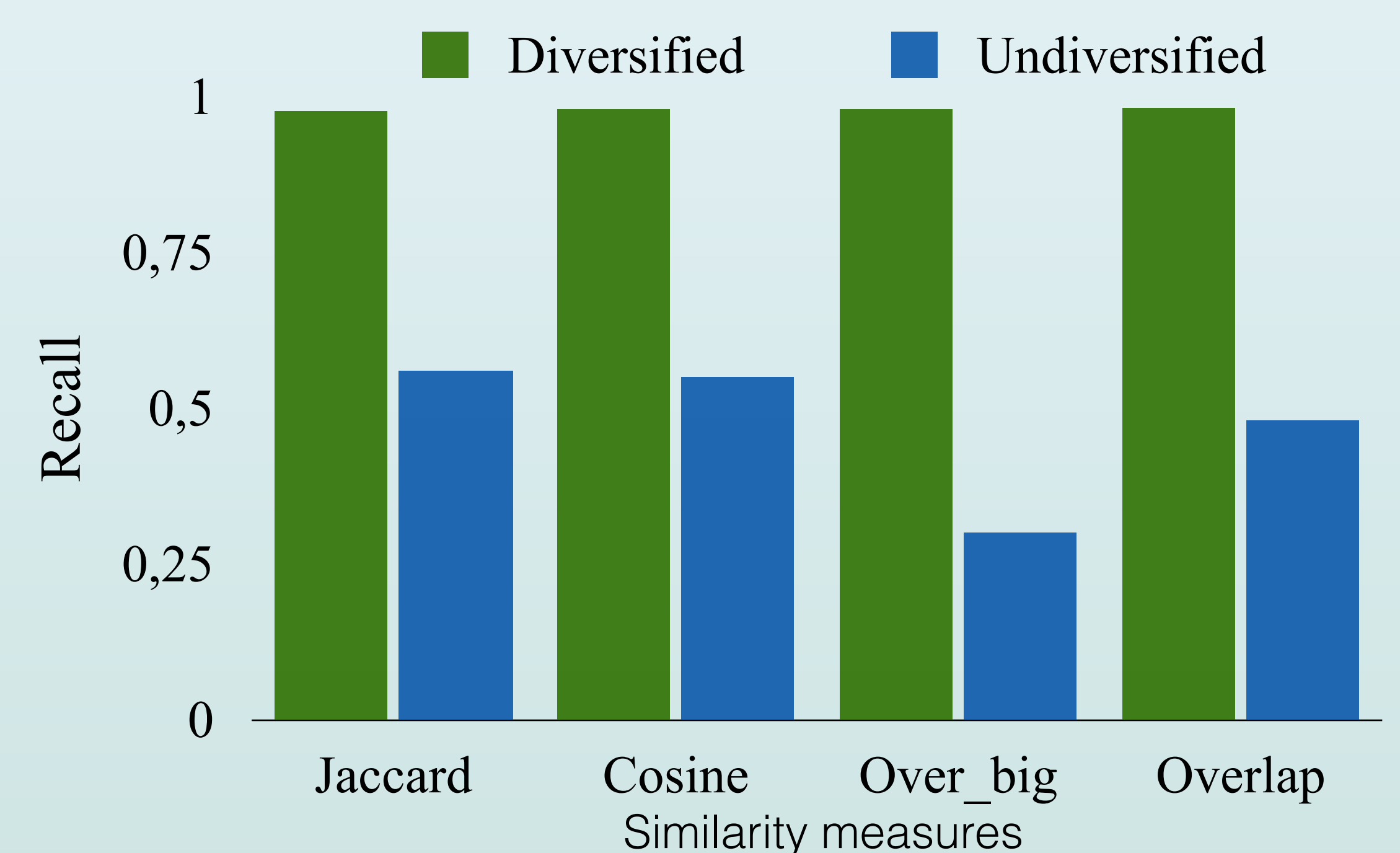
Recall: fraction of items that have been successfully recommended compared to a centralized implementation of the same algorithm.

[1] M. Servajean, E. Pacitti, S. Amer-Yahia, and P. Neveu, "Profile Diversity in Search and Recommendation" SRS'13.

[2] M. Servajean, E. Pacitti, M. Liroz-Gistau, S. Amer-Yahia, and A. El Abbadi, "Exploiting Diversification in Gossip-Based Recommendation" Globe'14.

[3] M. Servajean, E. Pacitti, M. Liroz-Gistau, A. Joly and J. Champ, "PlantRT: Multi-Site Diversified Search and Recommendation for Citizen Sciences" BDA'14 (submitted).