



HAL
open science

Introducing Cross Country News

Benjamin Renoust, Sébastien Poullot

► **To cite this version:**

Benjamin Renoust, Sébastien Poullot. Introducing Cross Country News: Bringing Big Data to Social Sciences. Global Young Scientist Summit, 2015, Singapore, Singapore. hal-01454983

HAL Id: hal-01454983

<https://hal.science/hal-01454983>

Submitted on 3 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Introducing Cross Country News Bringing Big Data to Social Sciences

Renoust Benjamin and Poullot Sébastien

(Japanese-French Laboratory for Informatics, National Institute of Informatics)

contact: renoust@nii.ac.jp

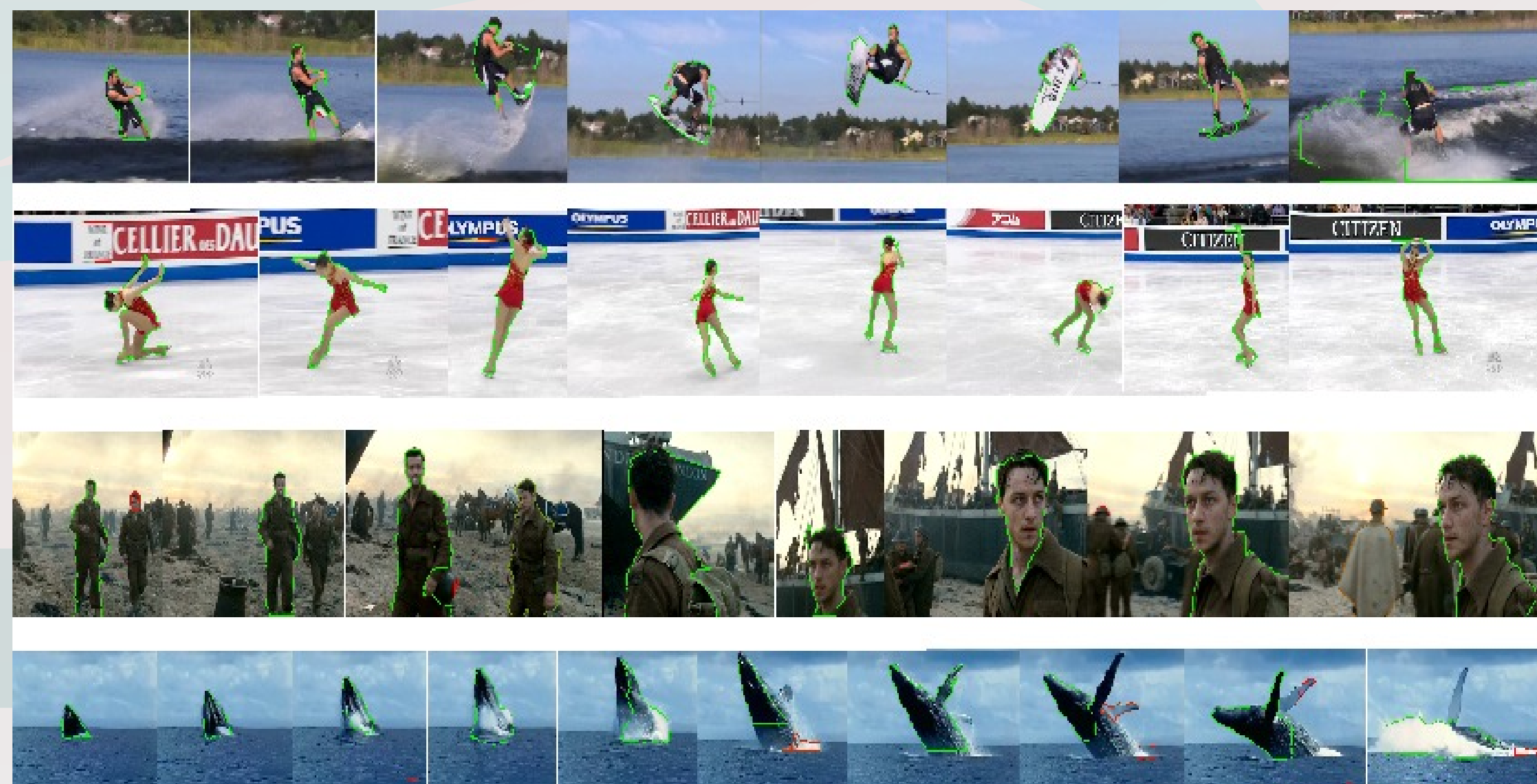


This project presents a multidisciplinary approach to Social Sciences and BigData analytics. It aims at observing the multimedia news across different diffusion sources (TV, newspapers, websites, social networks, etc.) and across different countries to compare them at the highest point of views. Our current capabilities are currently based upon the Japanese MEXT project MultiMedia Web Analysis Framework towards Development of Social Analysis Software (MMWeb), and the French ANR Project the TransMedia Observatory. We combine and advance techniques in multiple domains in close collaboration between computer scientists and social scientists to enable analysis. We contribute in three main disciplines, image analysis, information retrieval and analysis, and visual analytics, and enable answering a new set of social and societal questions.



Video and Image analysis

- Languages can be different, but the images remain the same. Finding and tracking an event can be done by finding emergent clusters of images.
- Images are used to illustrate news topics, but who actually issues the images? Is it issued by journalists of the main international press agencies? Is it issued by local journalists? Or does the user generated content grows more and more as a source for news?
- Faces and frequent objects like company logos are direct indicators of the occupation of the information landscape. Who are the main entertainers in each country? The main political figures that are represented in the other countries? Which brands are the most represented?



Information retrieval and analysis

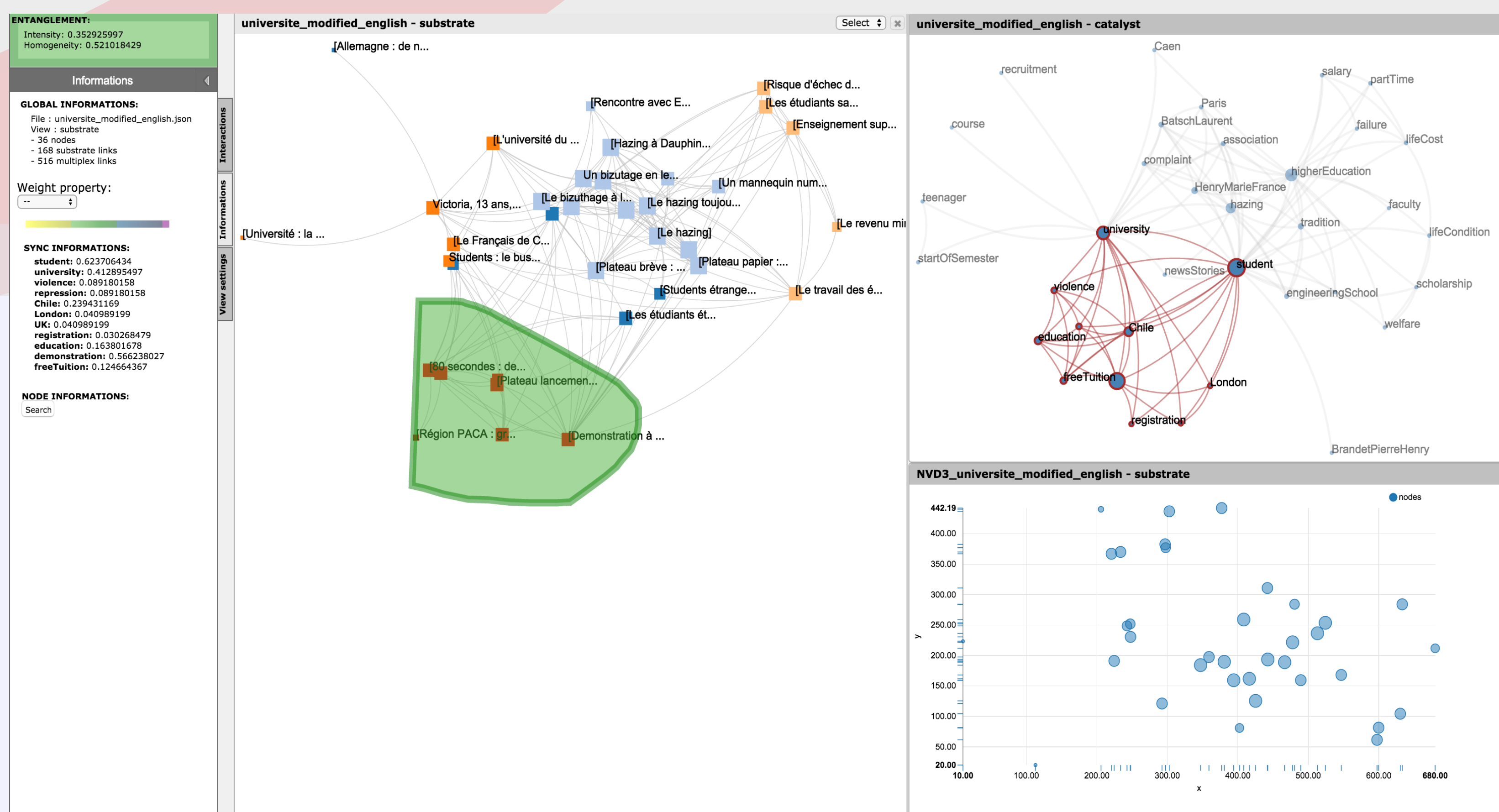
- We need to study multidimensional data and couple our extraction with external databases so we propose a complete analysis.
- For example, we want to identify the named entities referred to in the news such as *people* and *places*. This enables the study of many social networks, and ask questions, such as *which political figures often appear together?*
- Audience ratings can enable us to identify what interests the people. Frequency analysis enables to identify real news events as opposed to permanent background noises. For example, every Tuesdays in France, news are presenting the cinema movies coming out the next day.
- Geography based questions are also of prime interest since the proportion of local events can vary between countries, and also to consider when/on which topic a local event can cross its regional boundaries and even become international.



Visual Analytics

The presentation of the results are better integrated in a visual analytics framework. Visual representation and query allows close collaboration across disciplines. The visualization is designed with domain experts in order to display the right answers and the interaction to enable querying the right questions, making the whole BigData analysis transparent.

Network visual analytics especially allows very insightful observation of the data content, by focusing on the relationships between entities while displaying its general shape.



(1)OTMedia: the French TransMedia news observatory, Hervé, Nicolas, Viard Marie-Luce et al, 2013, ACMMM (2)Image Flows Visualization for Inter-media Comparison, Masahiko Itoh et al. PacificVis 2014 (3) Fouille au Corps des Media Français: Un exemple concret de Fouille Multimodale Transmedia, ML Viard, et al. 2014 (4) Measuring group cohesion in document collections, Renoust et al. ACMWI 2013