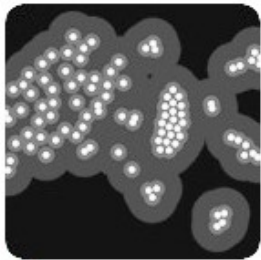


# **E-Diasporas**

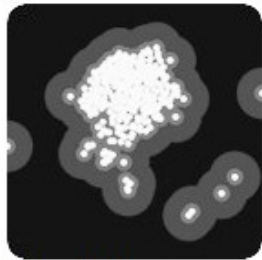
> International Forum on Migration Statistics 2018 – Session Migration Traceability

Dana Diminescu, Télécom ParisTech, I3-CNRS  
Mathieu Jacomy, SciencesPo Paris, Medialab  
Quentin Lobbé, Télécom ParisTech, LTCl, Inria Paris

## What is e-Diasporas **Atlas** ?



Lebanese corpus



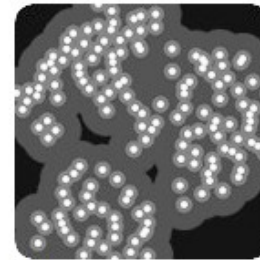
Macedonian corpus



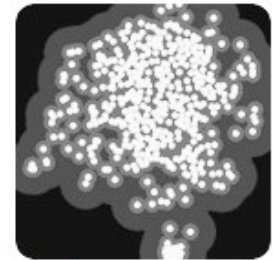
Mexican corpus



Moroccans on FB



Moroccan corpus



Nepali corpus

The *e-Diasporas Atlas* is above all a basic inventory of the networks on the Web represented in the form of graphs and analyzed.

## E-Diasporas project in **dates**

- > **2005-2008** – Development of Navicrowler & Gephi  
*by Mathieu Jacomy, Mathieu Bastien et Sebastien Heymann*
- > **2008**- The e-diasporas Atlas selected by  
the French Research National Agency (ANR)
- > **2012** – End of E-Diasporas
- > first runner up of its category visualization  
for the **2012** Digital Humanities Awards
- > **30** sections completed, a paper version, a iTunes App,  
a website, a distributed platform, an archive, a special issue  
of Social Science Information (Sage)

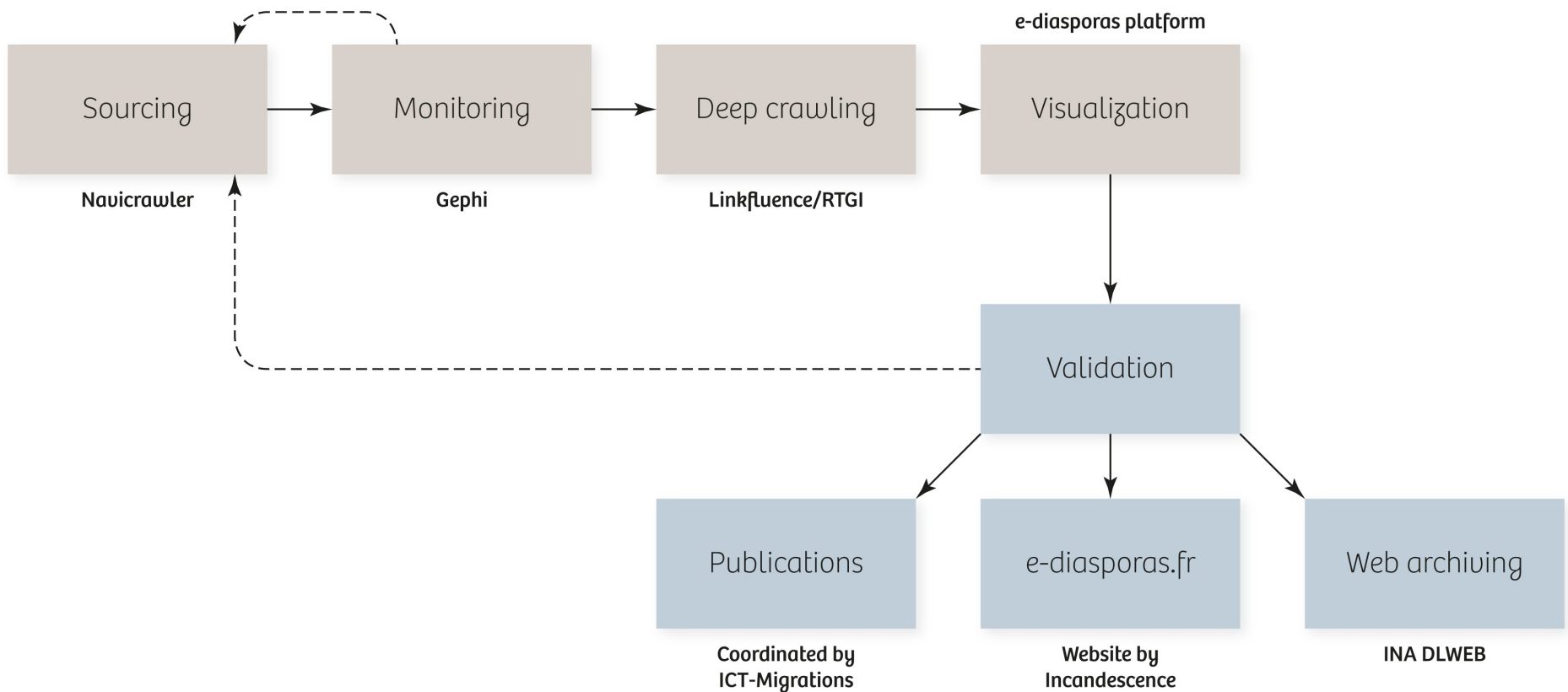
## E-Diasporas project in numbers

- > more than 9000 sites harvested and 70To of data
- > more than 80 researchers, engineers and designers of various institutions and careers
- > 11 Phd thesis defended
- > Gephi downloaded 100.000 times / month
- > more than 25.000 hours of works (very time-consuming)
- > more than 20 institutions involved

# Definition of an e-Diaspora

- > What we call *e-diaspora* is a **migrant collective** that organises itself and is active first and foremost on the Web: its practices are those of a community whose interactions are 'enhanced' by digital exchange.
- > An e-diaspora is also a **dispersed collective**, a heterogeneous entity whose existence rests on the elaboration of a common direction, a direction not defined once and for all but which is constantly renegotiated as the collective evolves.
- > An e-diaspora is an **unstable collective** because it is redrawn by every newcomer. It is self-defined, as it grows or diminishes not by inclusion or exclusion of members, but through a voluntary process of individuals joining or leaving the collective – simply by establishing hyperlinks or removing them from websites.

# The methodological chain



# How to build the categories relevant to the analysis?

> Network visualization displays topological clusters

## Modularity clustering is force-directed layout

Andreas Noack

*Institute of Computer Science, Brandenburg University of Technology, 03013 Cottbus, Germany*

(Dated: December 9, 2008)

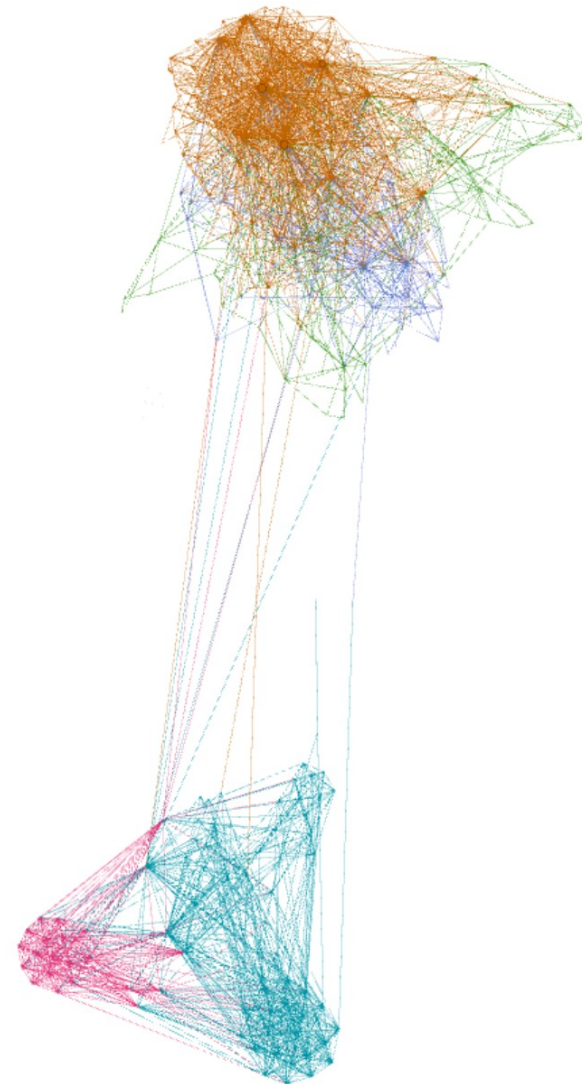
Two natural and widely used representations for the community structure of networks are clusterings, which partition the vertex set into disjoint subsets, and layouts, which assign the vertices to positions in a metric space. This paper unifies prominent characterizations of layout quality and clustering quality, by showing that energy models of pairwise attraction and repulsion subsume Newman and Girvan's modularity measure. Layouts with optimal energy are relaxations of, and are thus consistent with, clusterings with optimal modularity, which is of practical relevance because both representations are complementary and often used together.

PACS numbers: 89.75.Hc, 02.10.Ox

- Force-directed layouts display clusters
- These clusters are modules in the sense of Newman
- Modularity is used as community detection

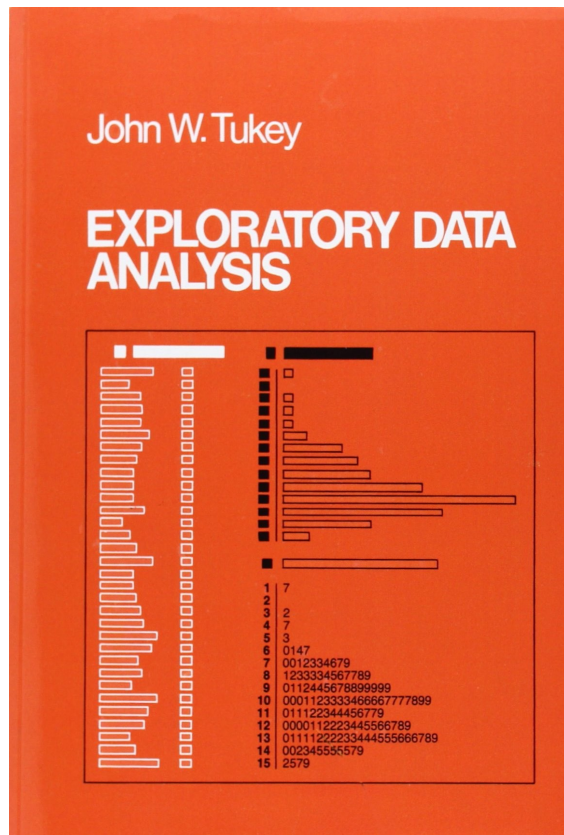
e-diasporas  
French Expatriates

Node color: Modularity  
Node position: Force-directed layout



# Generating hypotheses about topology

> Exploratory data analysis applied to the structure of hyperlinks



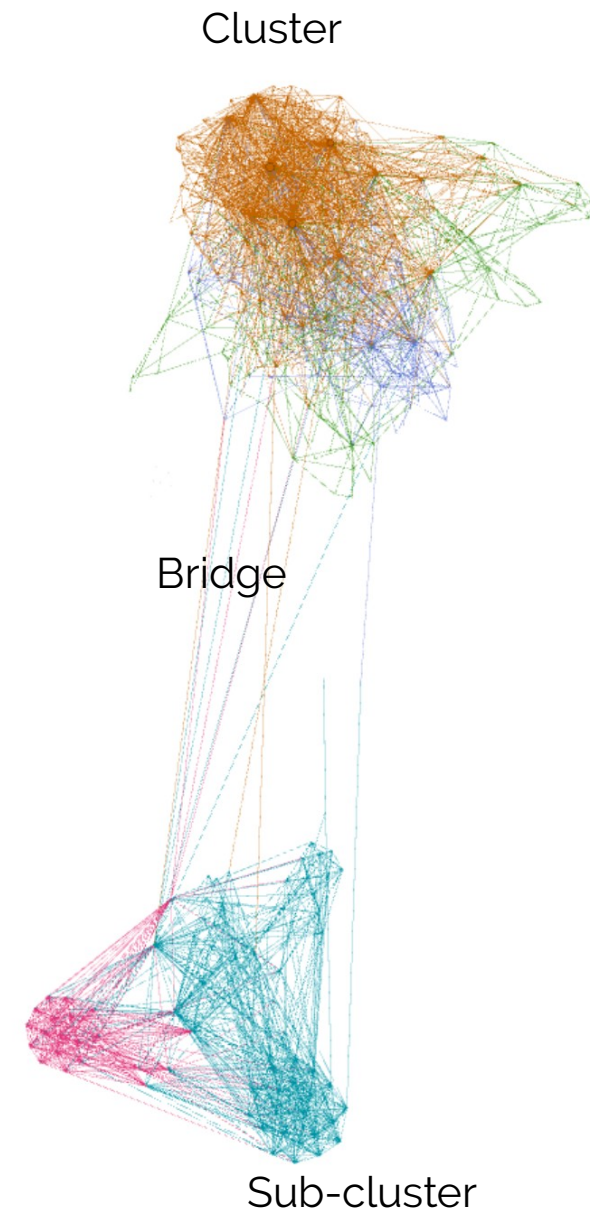
"The greatest value of a picture is when it forces us to notice what we never expected to see."

"Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."

—John W. Tukey

e-diasporas  
French Expatriates

Node color: Modularity  
Node position: Force-directed layout





## Categories are a compromise between topology and content

- > Researchers iterate to find the relevant categories



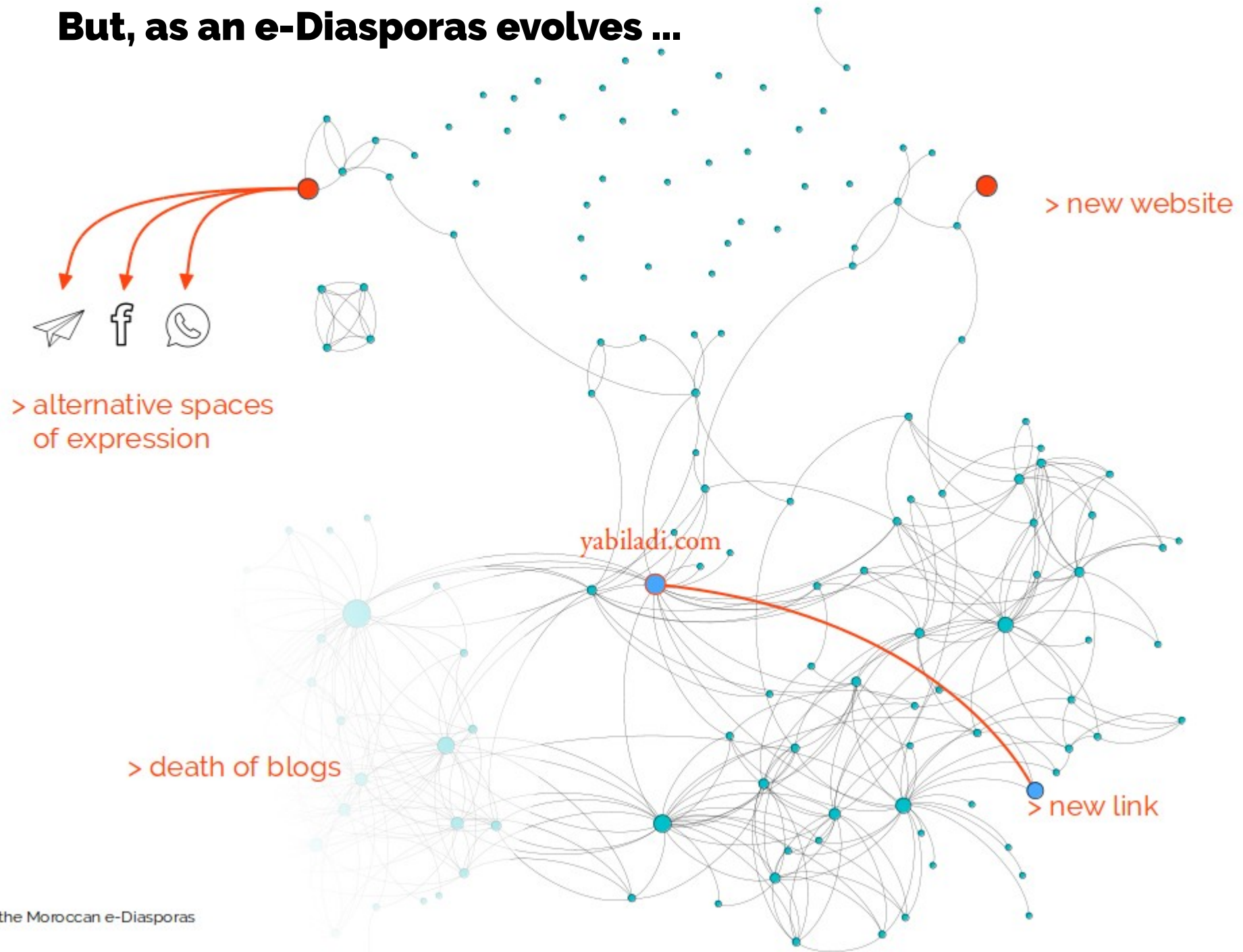
Topology and content challenge *a priori* categories. The researcher challenges clustering algorithms.

Different clusters may be  
the same community  
...but then why do they separate ?  
(ex: linguistic divides)

A single cluster may host  
different categories  
...but then why are they linked?  
(ex: alliance)

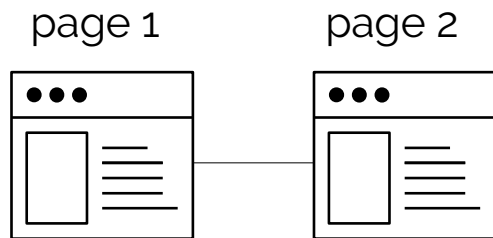
Some categories do not cluster at all  
(ex: competition)

## But, as an e-Diasporas evolves ...



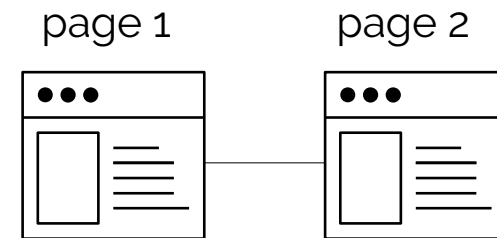
## ... it was decided to start archiving the corpora

> To keep a trace of the evolutions of every website



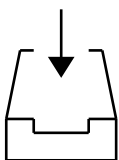
> time 1

record 1



> time 2

record 2



1030 M of webpages

70 TB

Crawled weekly or monthly, from April 2010 to September 2014

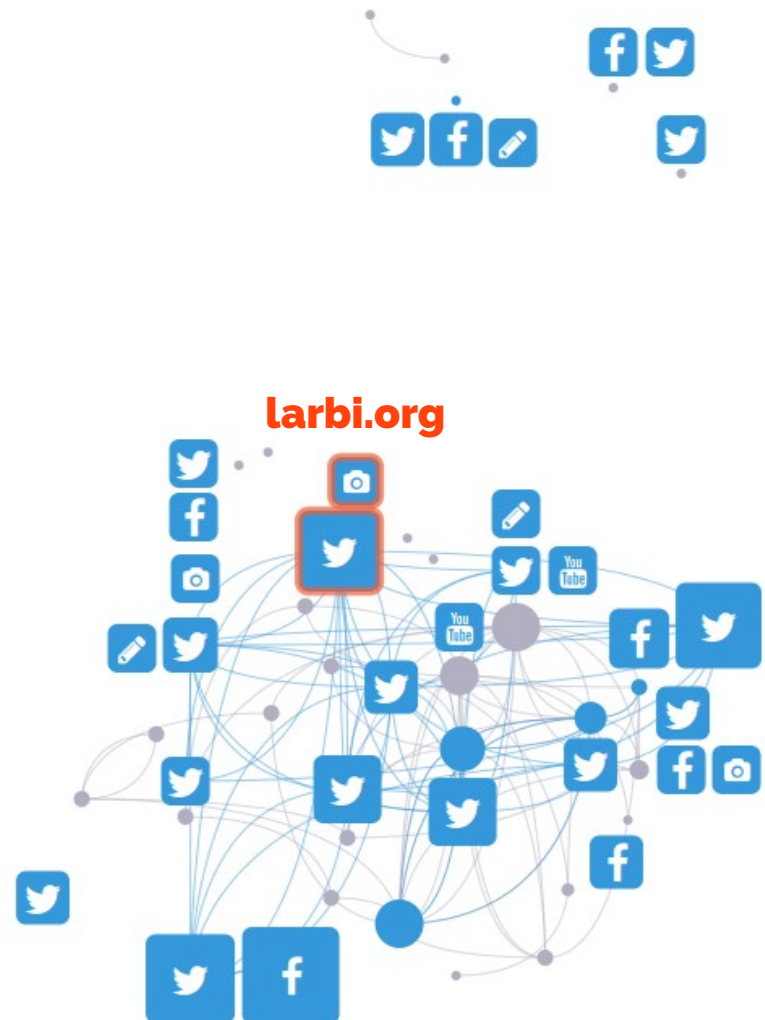
Hosted and performed by the INA

# What happened to the Moroccan blogosphere ?

> 2008 – 2010 : 48 blogs alive



> 2018 : 20 dead blogs, 23 deserted, 5 alive



# The blogosphere is continuously evolving

> From a single network to multiple networks inside multiple dedicated spaces

## larbi.org

2002 → 2013 : dead

- First social network linked in 2011 following the riots in Morocco in 2011
- Wrote a farewell message
- Then followed by some of its past commentators : 647 out of 4177
- Has a Twitter account since 2008
- Followed by 24300 people from Morocco, France, US, GB, Tunisia, Algeria, Egypt ...
- Has a images sharing account

blogosphere actors  
have emigrated to new web  
territories

2018's web is more about  
facing a flow of news than  
navigating through hyperlinks

a social network is a  
communication tool engaging  
a specific community