A Mesoscale Description of Networks' Dynamics Through Continuous Partitioning

Revealing the Structural Flow in Dynamic Networks

Pablo Jensen
IXXI, ENS Lyon, LabPhys, UMR 5672

Matteo Morini
IXXI, ENS Lyon, LIP, INRIA, UMR 5668

Márton Karsai
IXXI, ENS Lyon, LIP, INRIA, UMR 5668

Eric Fleury
IXXI, ENS Lyon, LIP, INRIA, UMR 5668

Patrick Flandrin
IXXI, ENS Lyon, LabPhys, UMR 5672

Large graphs require effective methods giving an appropriate **mesoscopic** description. Several approaches exist today to partition (static) graphs into communities (Fortunato, 2010). However, many networks are intrinsically dynamical, and describing them as static networks would cause loss of important informtion (Holme and Saramaki, 2012; Holme 2015). For example, dynamic processes such as the emergence of new scientific disciplines, their fusion, split or death need a mesoscopic description of the evolving network of scientific articles.

STATE OF THE ART

Methods to describe **evolving** networks from **static** networks:

- **aggregated network structure** → LOSS OF TEMPORAL INFORMATION
  Berger-Wolf and Saia, 2006

- **temporal time slices** w/INDEPENDENT structure for each slice, then connected → FUZZINESS – artifacts
  G. Palla et al, 2007, Rosvall and Bergstrom, 2010, Chavalarias and Cointet, 2013

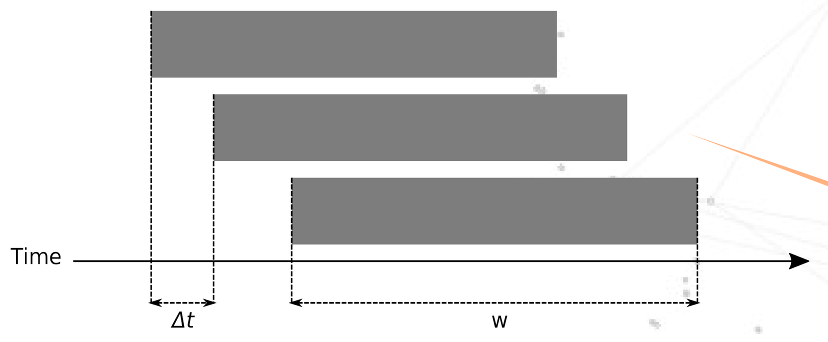
- Other existing approaches:
  Gauvin et al. 2015, Peel and Clauset, 2014, Mucha et al, 2010, Kawadia and Sreenivasan, 2012

Dataset and basic treatment

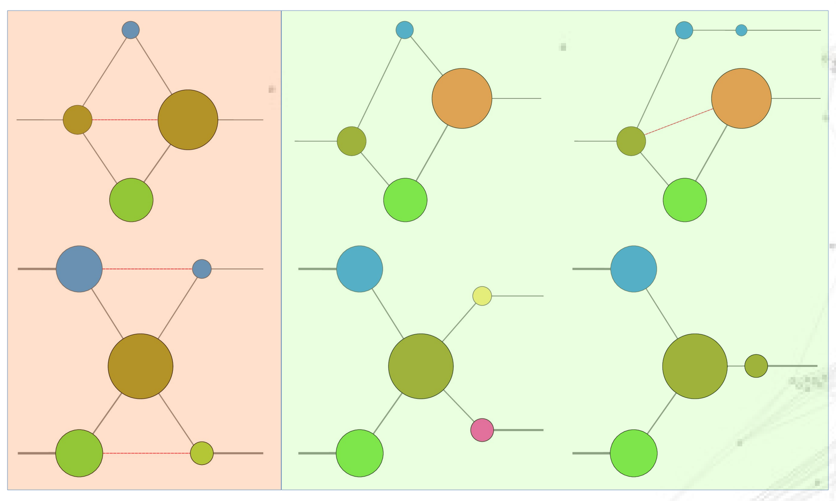
- 83 *key actors* (authors) since the early developments, 1970-2012
  → 6.500 WoS records (articles)

- Interesting facts:
  - bridges between Maths, Phys, Electr Eng
  - seminal paper: 1984
  - we track the stories of authors even *backwards* (when they didn't know they were going to work on wavelets in their future)

- Networks are made by **nodes** (articles) and **links** (shared references: Bibliographic Coupling w/at least 2, to limit noise)



$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$



BAD GOOD

Methodology

(1) Dataset divided into temporal windows of ***w*** (4) years, translated by ***Δt*** (1) years

(2) Community detection – **independent** for each window → structure that follows as closely as possible the **real mesoscale** dynamics, at the price of some noise. To selectively delete the noise, while keeping the real evolutions, one has to split or merge communities at each slice, depending on the relations between the successive communities on longer time scales

(3) **Similarities** (Jaccard Index) between communities in ***t*** vs. *t-2*, *t-1*, *t+1* and *t+2* are computed; the most important → *ancestor*, *predecessor*, *successor*, *grandchild*, strong long-term links.

"**Structural flows**" discern real evolutions from noise. Continuity and stability of social evolutions are kept into account on sensible time scales; e.g., a new scientific field does not appear and disappear in a single year).

(4) Long-term information → iteratively select all the time windows, and:

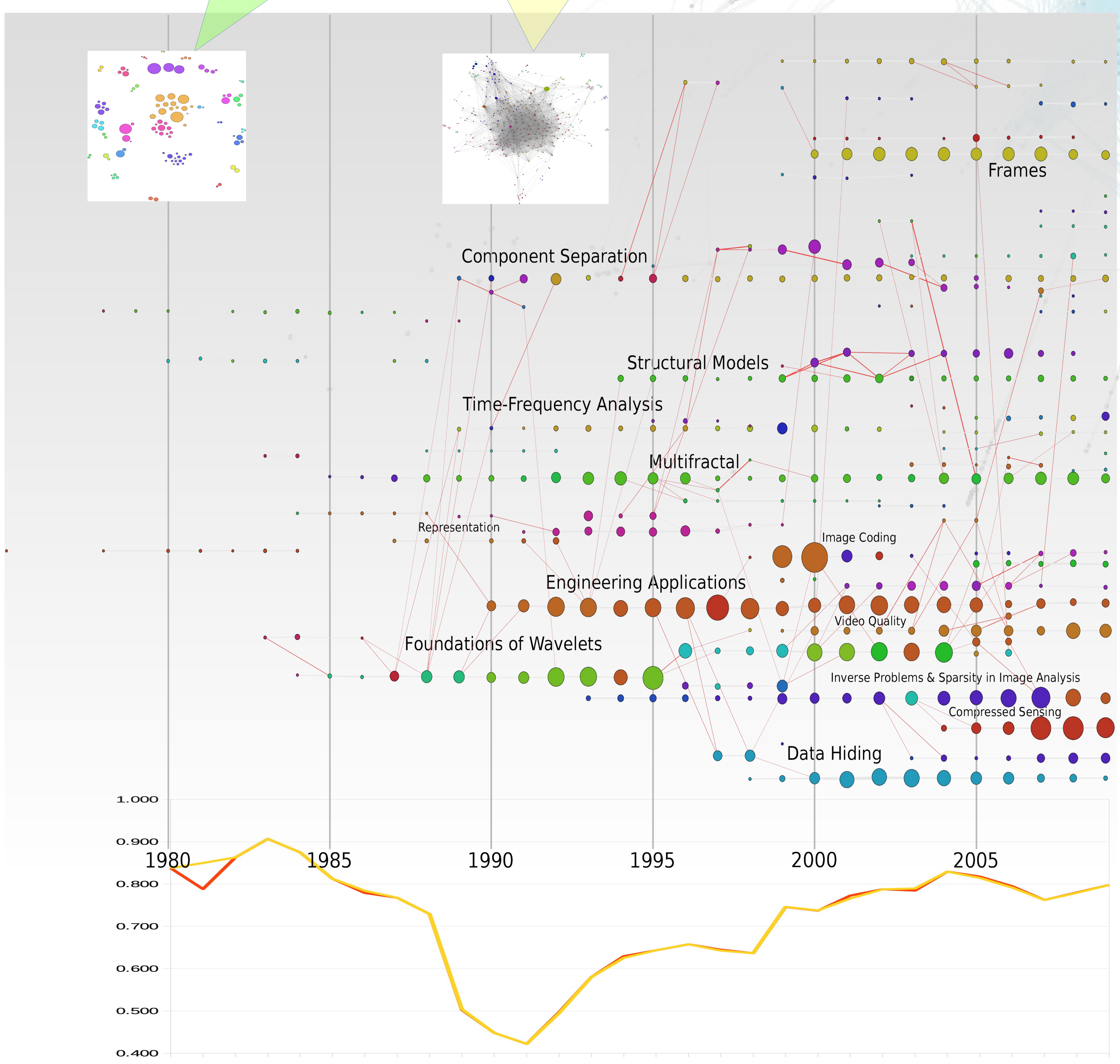
(4.a) **split** communities that have been wrongly merged by the time-independent community detection procedure, that is, where two streams appear independent but for one point.

(4.b) **merge** communities that appear to be artificial splits, that is where it is apparent that unduly split communities belong to a single stream.

At the end of the procedure, we obtain a description of the network evolution at the **mesoscale**, the unit of description being now several structural flows, i.e. streams of connected communities. Note however that the final description may depend on the set of initial partitions (more on this below). To render our method robust, we compute a "continuity" score (see below) for different final descriptions and use the one with the highest score, leading to the richer story that can be told avoiding noise. The merit of our approach is, by eliminating most of the noise, to limit these complex turbulent regions to the real transformations that should not be discarded : things should be made simple, but not too simple.

Well Structured Network, High Modularity (1980)

Poorly Structured Network, Low Modularity (1991)



Modularity dips to ~0.42 in 1991: mixing of different disciplinary traditions around *wavelets;* negligible difference in modularity *before (Q)* and *after* (Q_post) filtering.

Continuity score: Cs

A quality measure, "Cs", can be computed for every historical graph, both before and after applying the filtering procedure described under "Methodology", at step (4). Filtering does not worsen the Cs by construction: *Cs_post ≥ Cs_pre*.

The positive (+) component comes from "structural" events, persistent splits and merges; the negative (-) consists of ephemeral events; both are normalized by the total node size.

We select the best history in terms of Cs, out of the population resulting from different initial Louvain partitions.

$$C_S = \frac{\overbrace{\sum_{u \in (u_s \cup u_m)} s_u}^{+} - \overbrace{\sum_{u \in (u_r \cup u_x)} s_u}^{-}}{\sum_{u \in G} s_u}$$

Where:
$u = \forall u \in G$, all nodes.
s_u = Size of node u
u_s = In all structural split occurrences; resulting split nodes
u_m = In all structural merge occurrences; resulting merged node
u_r = In all ephemereal split occurrences; resulting split nodes
u_x = In all ephemereal merge occurrences; resulting merged nodes

Results and storytelling

High modularity (isolated clusters), low peak (~1991) of Q, all mixed up, Q increases again

The analysis shows that there are three main stages. In an initial phase (before ~ 1985), researchers work in different, relatively unrelated fields and modularity is high. Then, around 1991, wavelets appear as a common topic whose use gains momentum, defining a new, specific field that interlinks scholars, leading to a minimum in modularity. After this, modularity increases again, pointing to a new, softer divergence, as the initial levels are not reached. Wavelets become a mature tool, that are less an object of interest per se, serving instead a more ancillary role within specialized communities and paving the way for new avenues of research, by developing new tools (as exemplified by the emergence of "compressed sensing") or applying wavelets to specific, relatively unrelated, domains.

Concluding remarks and future work

To make sense of transformations, we need structural flows, i.e. evolving categories that can, at the same time, readily adapt to the changes and maintain the continuity of the description. Our method starts from the idea that the unity of an evolving social process rests on the continuity of its transformations, and uses the available mid-term temporal information to reveal **structural trends** from noisy data, **without** the assumption of an **a priori community structure**. It can handle naturally changes in the number of nodes, and be adapted to any partitioning method and to any similarity measure between communities at different times. Used on scientific data, our method automatically produces a rich historical account, an objective raw material to be discussed by science historians.

There is much room for improvement. The relevant time scales (w, delta t) have to be chosen from expert knowledge, and we cannot deal with real-time data, as we use the future to infer the best present partition. We now work to introduce, through a hidden Markov model, an explicit meso temporal scale at which transformations (splits/merges) are supposed to happen for a pair of streams.

INVENTEURS DU MONDE NUMÉRIQUE
Institut des systèmes complexes Complex Systems Institute Rhône-Alpes
IXXI
ENS DE LYON