Academic team formation
000

Framework and Protocol
00000

Results
0000

Conclusion & Perspectives
00

SUNBELT XXX

————

Evolving hypergraphs to appraise

academic team formation processes

Carla TARAMASCO, Jean-Philippe COINTET and Camille ROTH

CREA (CNRS), INRA-SenS (INRA), CAMS (EHESS, CNRS) & ISCPIF

July 2, 2010

Mechanisms of academic collaboration

Academic collaboration - a long tradition of research :

- collaborative activity growth (*M. Smith*, 1958)
- International collaboration growth (*Wagner, C. S., Leydesdorff, L.*, 2005)
- co-authorship network as a complex evolving networks, (*Moody*, 2004 ;), (*Newman, M. E. J.*, 2004)

Sources of Collaboration, it's all about proximity

- Spatial/Physical (*Kraut, R.E., Fussell, S.R., Brennan, S.E., Siegel, J.*, 2002), (*Katz*, 2002)
- Social distance (*WO Hagstrom,* 1965)
- Intellectual (*Cowan, R., Jonard, N., Zimmermann, J.-B*, 2002)

Academic team formation
○●○

Framework and Protocol
○○○○○

Results
○○○○
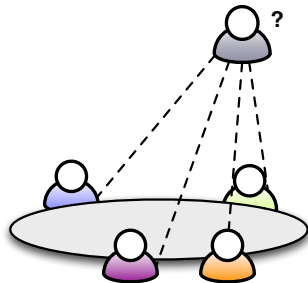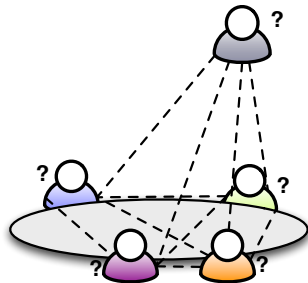
Conclusion & Perspectives
○○

The Team Level & Networks

Limits when focusing on the level of individual

- only dyads, overlook the influence of characteristics expressable at the mesolevel of the team itself,

- team formation processes ≠ sum of individual rationalities.

Toward Meso-level approach

- focus on *teams* rather than pairs of agents interacting together,

Academic team formation
○●○

Framework and Protocol
○○○○○

Results
○○○○

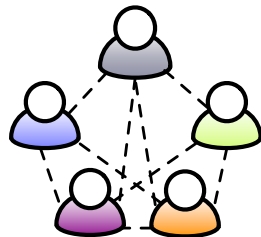Conclusion & Perspectives
○○

The Team Level & Networks

Limits when focusing on the level of individual

- only dyads, overlook the influence of characteristics expressable at the mesolevel of the team itself,

- team formation processes $\neq$ sum of individual rationalities.

Toward Meso-level approach

- focus on *teams* rather than pairs of agents interacting together,

- hypergraphs not cliques.

Academic team formation
○●○

Framework and Protocol
○○○○○

Results
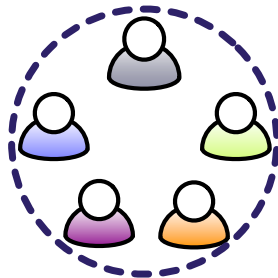○○○○

Conclusion & Perspectives
○○

The Team Level & Networks

Limits when focusing on the level of individual

- only dyads, overlook the influence of characteristics expressable at the mesolevel of the team itself,

- team formation processes $\neq$ sum of individual rationalities.

Toward Meso-level approach

- focus on *teams* rather than pairs of agents interacting together,

- hypergraphs not cliques.

Academic team formation
○●○

Framework and Protocol
○○○○○

Results
○○○○

Conclusion & Perspectives
○○

The Team Level & Networks

Limits when focusing on the level of individual

- only dyads, overlook the influence of characteristics expressable at the mesolevel of the team itself,

- team formation processes ≠ sum of individual rationalities.

Toward Meso-level approach

- focus on *teams* rather than pairs of agents interacting together,
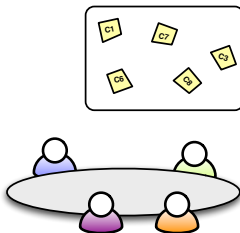
- hypergraphs not cliques.

Hybrid Networks : Actors and Concepts

Collaboration also depends on cognitve properties

epistemic dynamics $=$ reconfiguration of collectives made of :

- actors,
- concepts.

Question :

How new teams are formed given both *social* and *conceptual* past acquaintances of scientists ?

Academic team formation
000

Framework and Protocol
●0000

Results
0000

Conclusion & Perspectives
00

Datasets

Experimental protocol

- for 4 different datasets describing research production over $\sim$ 20 years

Academic team formation
000

Framework and Protocol
●0000

Results
0000

Conclusion & Perspectives
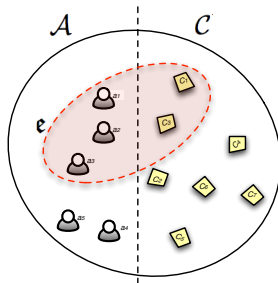00

Datasets

Experimental protocol

- for 4 different datasets describing research production over $\sim$ 20 years
- extract the set of agents $\mathcal{A}$ and a set of pertinent concepts $\mathcal{C}$
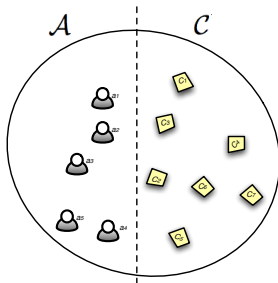
Datasets

Experimental protocol

- for 4 different datasets describing research production over $\sim$ 20 years
- extract the set of agents $\mathcal{A}$ and a set of pertinent concepts $\mathcal{C}$
- each paper is defined as a hyperlink : $\mathfrak{e} \in \mathcal{P}(\mathcal{A} \cup \mathcal{C})$, that is the joint grouping of both agents and concepts

Datasets

Experimental protocol

- for 4 different datasets describing research production over $\sim$ 20 years
- extract the set of agents $\mathcal{A}$ and a set of pertinent concepts $\mathcal{C}$
- each paper is defined as a hyperlink : $\mathfrak{e} \in \mathcal{P}(\mathcal{A} \cup \mathcal{C})$, that is the joint grouping of both agents and concepts
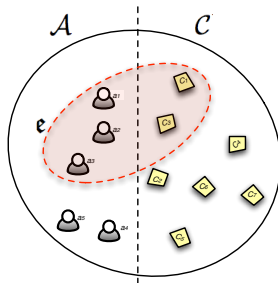
Projection operator

One can decompose an hyperlink $\mathfrak{e}$ on any subset of $\mathcal{A} \cup \mathcal{C}$ with operator $\cdot$
Especially the set of co-authors of article $\mathfrak{e}$ is given by :
$\mathfrak{e}^{\mathcal{A}} = \mathfrak{e} \cap \mathcal{A} : \{a_1, a_2, a_3\}$,
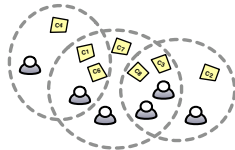its concepts are defined as :
$\mathfrak{e}^{\mathcal{C}} = \mathfrak{e} \cap \mathcal{C} : \{c_1, c_3\}$ in this example

Hybrid Networks : Actors and Concepts

Epistemic Hypergraph

- *epistemic hypergraph* = triple $(\mathcal{A}, \mathcal{C}, \mathfrak{E})$, where $\mathfrak{E} \subseteq \mathcal{P}(\mathcal{A} \cup \mathcal{C})$
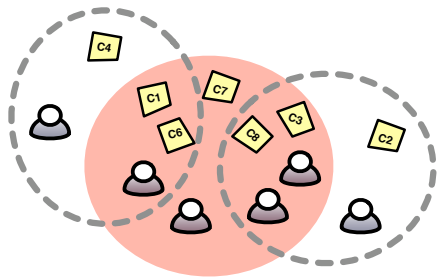- The epistemic hypergraphs is growing with time : $\mathfrak{E}_t$

$\mathfrak{E}_{t-1}$

$\cup \quad \Delta\mathfrak{E}_t$

$= \quad \mathfrak{E}_t$

Academic team formation
000

Framework and Protocol
00●00

Results
0000

Conclusion & Perspectives
00
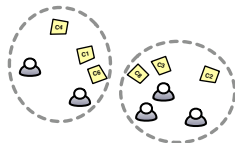
Definitions

Homogeneity of teams and expertise ratio

- neophytes *vs* experts
- $\xi_c(\mathfrak{e})$ *expertise ratio of an article $\mathfrak{e}$ given a concept $c \in \mathfrak{e}^{\mathcal{C}}$:*

$$\xi_c(\mathfrak{e}) = \frac{|\{a \in \mathfrak{e}^{\mathcal{A}} \mid a \text{ expert in } c\}|}{|\{a \in \mathfrak{e}^{\mathcal{A}}\}|}$$

examples :

$\xi_{c_1}(\mathfrak{e}) = 2/5$
$\xi_{c_6}(\mathfrak{e}) = 2/5$
$\xi_{c_7}(\mathfrak{e}) = 0$, etc.
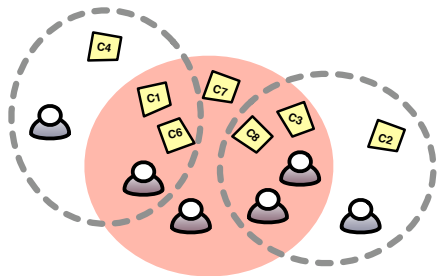
Definitions

Hypergraphic repetition

- Originality of the composition of a team : *social originality* and *conceptual originality*

- Set of nodes repetition : is there at least one previously existing hyperlink including this set ?

$$\rho_t(\mathfrak{e}) = \begin{cases} 1 & \text{if } \exists \mathfrak{e}' \in \mathfrak{E}_{t-1}, \mathfrak{e} \subseteq \mathfrak{e}' \\ 0 & \text{otherwise.} \end{cases}$$

- Hypergraphic repetition = proportion of subsets of $\mathfrak{e}$ that are repeated :

$$r_t(\mathfrak{e}) = \frac{1}{2^{|\mathfrak{e}|} - |\mathfrak{e}| - 1} \sum_{\substack{\mathfrak{e}' \subseteq \mathfrak{e} \\ |\mathfrak{e}'| \geq 2}} \rho_t(\mathfrak{e}')$$

examples :

social hypergraphic repetition rate $r_t(\mathfrak{e}^{\mathcal{A}}) = \frac{1}{4}$

conceptual hypergraphic repetition rate $r_t(\mathfrak{e}^{\mathcal{C}}) = \frac{2}{11}$

Academic team formation
000

Framework and Protocol
0000●

Results
0000

Conclusion & Perspectives
00

Estimating Propensities of team formation

Null-model of hypergraph

we generate at each time step a set of new teams $\widetilde{\Delta\mathfrak{E}_t}$ which respects the following distributions :
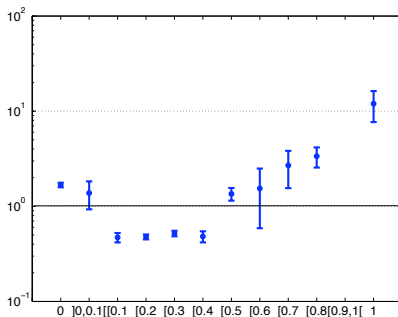
- same distribution of sizes of new hyperlinks (same dist. $|\mathfrak{e}^{\mathcal{A}}|$ and $|\mathfrak{e}^{\mathcal{C}}|$ for $\mathfrak{e} \in \Delta\mathfrak{E}_t$)
- same distribution of participations of elements in these new hyperlinks.

Propensity

Given a measure f (*e.g.* hypergraphic repetition) on a hyperlink, we compute the likeliness that for a new team $\mathfrak{e}$, $f(\mathfrak{e}) = x$.

$$\Pi_t(x) = \frac{\left|\{\mathfrak{e} \in \Delta\mathfrak{E}_t \text{ such that } f(\mathfrak{e}) = x\}\right|}{\left|\{\mathfrak{e} \in \widetilde{\Delta\mathfrak{E}_t} \text{ such that } f(\mathfrak{e}) = x\}\right|}$$

Academic team formation
ooo

Framework and Protocol
ooooo

Results
●ooo

Conclusion & Perspectives
oo

Teams expertise ratio

The curve is U-shaped :

teams are more likely to be mainly composed with all-neophytes or all-experts,

mixed teams are less frequent than expected from our null-model

Propensity that team have a given expertise ratio
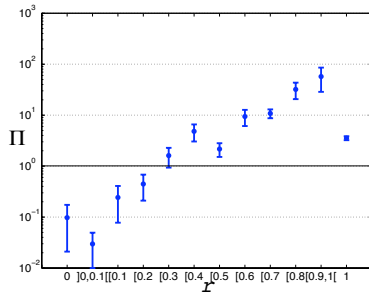
computed over 10 bins and shown on one dataset

Teams hypergraphic repetition rate

Teams hypergraphic repetition rate propensity

Likeliness to produce teams with a given social (*left*) and conceptual (*right*) hypergraphic rate of repetition (computed over 10 bins and shown on one dataset)
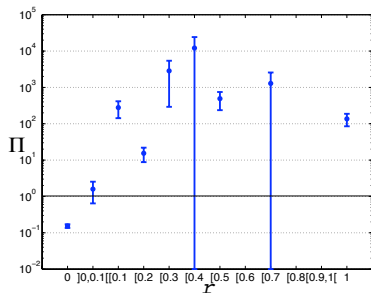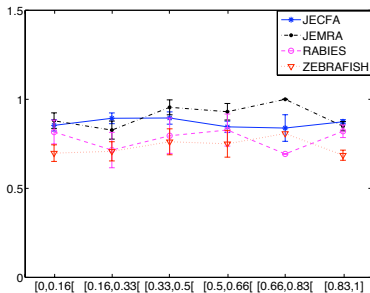

high proportion of interaction repetitions

bias towards gathering groups of concepts which were previously associated

Academic team formation
ooo

Framework and Protocol
ooooo

Results
oo●o

Conclusion & Perspectives
oo

Are hypergraphic repetition rates correlated ?

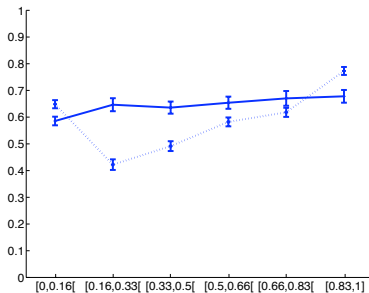We observe no correlation

contrarily to intuition, new semantic associations do not stem more from original teams than from repeated teams

Correlation between *agents repetition ratio* and *average semantic repetition ratio*

Average semantic hypergraphic repetition ratio (*y-axis*) for a given range of social hypergraphic repetition ratio (*x-axis*), computed on 6 bins and shown for every datasets

Academic team formation
ooo

Framework and Protocol
ooooo

Results
ooo●

Conclusion & Perspectives
oo

Are hypergraphic repetition rates correlated ?

We observe no correlation between expertise ratio and semantic originality

yet, expertise ratio is broadly growing with social repetition ratio

social originality is increased when there is a mixed proportion of experts, but not semantic originality

Correlation between *expertise ratio* and *hypergraphic repetition ratios*

Average hypergraphic repetition ratios (*y-axis*) with respect to expertise ratios (*x-axis*) : social (*dashed line*) and semantic (*plain line*) cases, computed on 6 bins and shown for one dataset
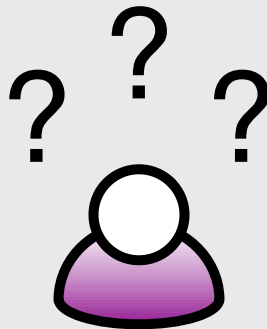
Conclusion & Perspectives

Strictly social and semantic associations

- formal framework to appraise the underpinnings of collaboration formation with a hypergraphic approach which encompasses both the meso-level of teams and the joint dynamics of social and semantic features.

- (i) high likeliness to repeat previous collaborations patterns, along with a polarization between groups made of experts only or made of non-experts only
  (ii) similarly, sensible semantic confinement where associations of concepts depend largely on the repetition of previous associations.
  (ii) However, the originality of a paper does not seem to stem from an original composition of the underlying team

Perspectives on models of academic collaboration

- In line with our results, it should also be possible to determine which features, at the level-team, favor better collaborations — not only in terms of semantic originality, but also in terms of quality and creativity of output

Academic team formation
000

Framework and Protocol
00000

Results
0000

Conclusion & Perspectives
0●

Questions

Reference

Academic team formation as evolving hypergraphs
Taramasco, C., Cointet, J.P. and Roth, C., Scientometrics, 2010