



Communities detection and analysis of their dynamics in collaborative networks

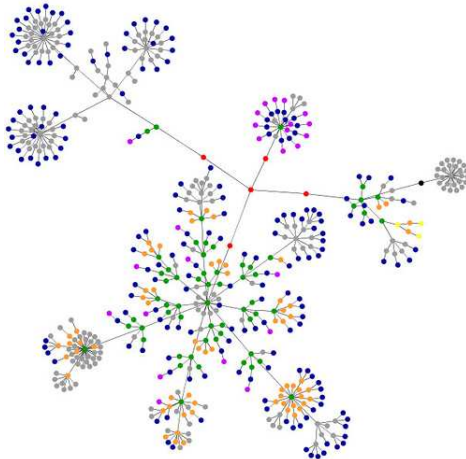
Manel Ben Jdidia, Céline Robardet and Eric Fleury

CITI - INSA-Lyon
LIRIS - INSA-Lyon
INRIA ARES - INSA-Lyon

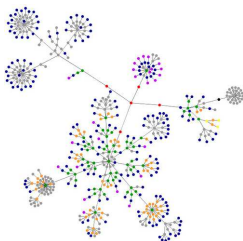
Outline

- 1 Introduction
 - Evolving graphs and collaborative networks
 - Infocom Data: a co-authorship network
- 2 Communities detection in evolving graphs
 - Identification of stable collaborations
 - Identification of evolving communities
- 3 Impact of the graph topology on the evolution of identified groups
 - Dependencies on the number of neighbors that belong to the group
 - Dependencies on a broader range of topology features
- 4 Conclusion and future work

Social network



Social network



Study of social networks:

- How pattern of interactions among social entities can originate social outcomes?

Study of the evolving of a social network:

- How the dynamics of interactions enables to characterize the global organization of the network?

Goals: identification and study of communities

Community

- Set of social entities which interact much more together than with outside entities
 - Social entities from the same community share similar characteristics: degree, neighborhood, ...
-
- Identification of stable communities
 - Identification of time evolving communities
 - Study of the impact of the graph topology on the evolution of a well identified community

Studying a co-authorship network

INFOCOM conference data

- 4030 published papers from 1985 to 2007
- 5164 authors

Co-authorship network characteristics

	arXiv	Infocom
n (nb. of vertices)	16 401	5 164
m (nb. of edges)	29 552	8 918
k (average degree)	3.60	3.45
δ (density)	2.2e-4	6.6e-4
d (average distance)	7.18	6.92
diameter	20	18

Studying a co-authorship network

INFOCOM conference data

- 4030 published papers from 1985 to 2007
- 5164 authors

Co-authorship network characteristics

	arXiv	Infocom
n (nb. of vertices)	16 401	5 164
m (nb. of edges)	29 552	8 918
k (average degree)	3.60	3.45
δ (density)	2.2e-4	6.6e-4
d (average distance)	7.18	6.92
diameter	20	18

Studying a co-authorship network

INFOCOM conference data

- 4030 published papers from 1985 to 2007
- 5164 authors

Co-authorship network characteristics

	arXiv	Infocom
n (nb. of vertices)	16 401	5 164
m (nb. of edges)	29 552	8 918
k (average degree)	3.60	3.45
δ (density)	2.2e-4	6.6e-4
d (average distance)	7.18	6.92
diameter	20	18

Studying a co-authorship network

INFOCOM conference data

- 4030 published papers from 1985 to 2007
- 5164 authors

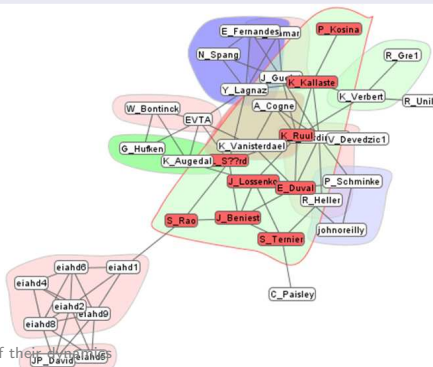
Co-authorship network characteristics

	arXiv	Infocom
n (nb. of vertices)	16 401	5 164
m (nb. of edges)	29 552	8 918
k (average degree)	3.60	3.45
δ (density)	2.2e-4	6.6e-4
d (average distance)	7.18	6.92
diameter	20	18

Communities detection in evolving graphs

Community definition

A partition $\mathcal{P} = \{\mathcal{C}_1, \dots, \mathcal{C}_k\}$ of the vertices of a graph $G = (V, E)$ represents a good community structure if the proportion of edges inside the \mathcal{C}_i (internal edges) is high compared to the proportion of edges between them.



State of the art

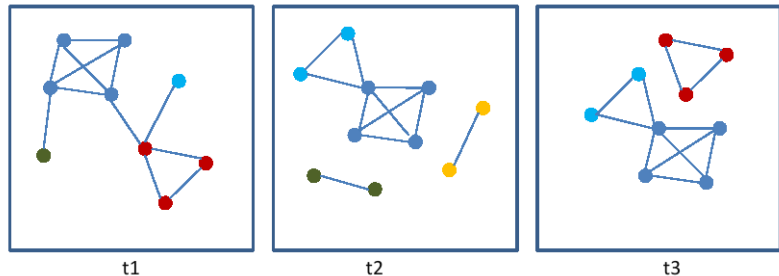
- Leicht et al. use the temporal evolution of the degree as descriptive feature of each node and use the EM clustering algorithm to identify communities.
- Hopcroft et al. propose to identify communities at each time points (using an agglomerative clustering algorithm). Then, they link clusters of two consecutive time points using a matching algorithm.

Identifying stable collaborations

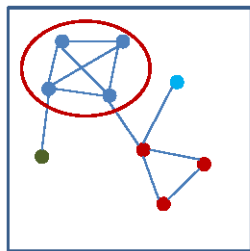
Identifying groups of authors that frequently co-sign papers

- We consider the co-author graphs $G_t = (V_t, E_t)$ for each year t such that $\{x, y\} \in E_t$ if $x, y \in V_t$ and x and y co-sign a paper published year t .
- Computing frequent maximal cliques in this set of graphs $\{G_t \mid t = 1985, \dots, 2007\}$.

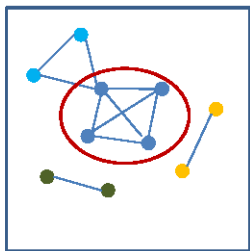
Groups of authors that frequently co-sign papers



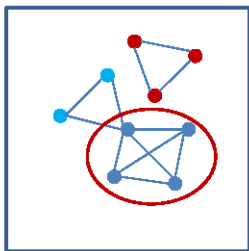
Groups of authors that frequently co-sign papers



t1

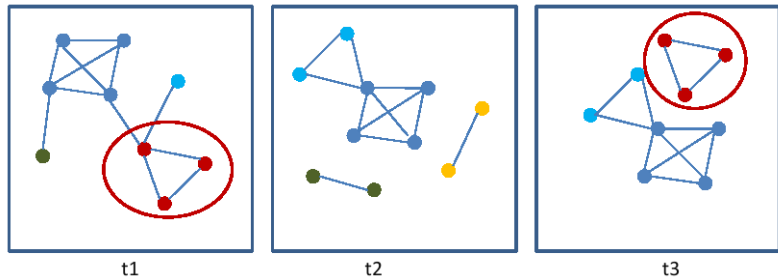


t2

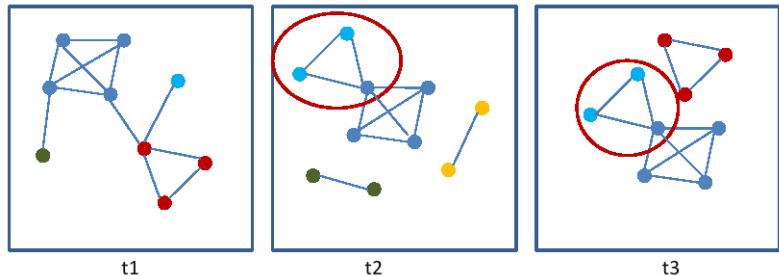


t3

Groups of authors that frequently co-sign papers



Groups of authors that frequently co-sign papers



Maximal cliques of at least 3 authors and 3 years

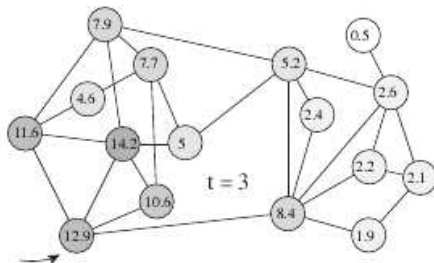
#	Author cliques C			Nb years
1	M. Ahamad (2)	M. Ammar (34)	J. Bernabeu-Auban (2)	3
2	C. Barnhart (2)	A. Ephremides (8)	J. Wieselthier (6)	3
3	D. Dutta (2)	A. Goel (19)	J. Heidemann (15)	3
4	L. Kalampoukas (4)	K. Ramakrishnan (23)	A. Varma (2)	3
5	R. Doverspike (7)	G. Li (6)	D. Wang (5)	3
6	M. Conti (4)	E. Gregori (4)	L. Lenzini (8)	3
7	S. Acharya (3)	B. Gupta (4)	P. Risbood (4)	3
8	M. Kodialam (16)	T. Lakshman (26)	S. Sengupta (6)	3
9	S. Low (30)	A. Tang (5)	J. Wang (8)	3
10	S. Donatelli (4)	M. Marsan (24)	F. Neri (18)	4
11	D. Figueiredo (7)	J. Kurose (53)	D. Towsley (88)	3
12	O. Frieder (6)	X. Li (10)	P. Wan (18)	3
13	Q. Fang (6)	J. Gao (10)	L. Guibas (6)	4
14	M. Azizoglu (8)	A. Somani (9)	S. Subramaniam (10)	3
15	G. Iannaccone (7)	S. Jaiswal (9)	J. Kurose (53)	3
16	Y. Breitbart (9)	M. Garofalakis (14)	R. Rastogi (18)	4
17	C. Hollot (9)	V. Misra (16)	D. Towsley (88)	3
18	I. Cidon (24)	A. Khamisy (7)	M. Sidi (26)	5
19	A. Bianco (7)	E. Leonardi (18)	M. Marsan (24)	4
20	A. Bianco (7)	E. Leonardi (18)	F. Neri (18)	5
21	S. Bhattacharjee (18)	K. Calvert (10)	E. Zegura (19)	3
22	P. Giaccone (9)	E. Leonardi (18)	F. Neri (18)	3
23	P. Giaccone (9)	E. Leonardi (18)	M. Marsan (24)	3
24	E. Leonardi (18)	M. Marsan (24)	M. Mellia (12)	4
25	E. Leonardi (18)	M. Marsan (24)	F. Neri (18)	8

Identifying evolving communities

The random walk approach

Discovering evolving communities with a random walk based method

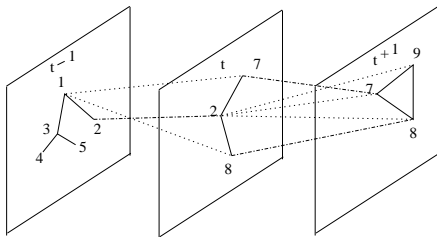
Random walks are trapped into dense area thanks to the high density of links in the community.



A single evolving graph

Dynamic communities

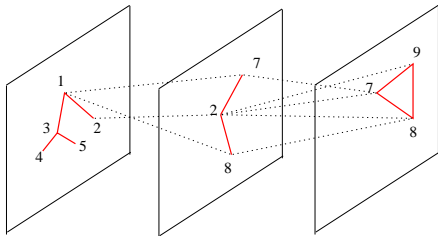
Finding communities that gather direct co-authors and also authors that share common co-authors.



A single evolving graph

Dynamic communities

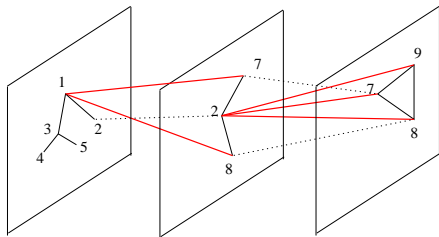
Finding communities that gather direct co-authors and also authors that share common co-authors.



A single evolving graph

Dynamic communities

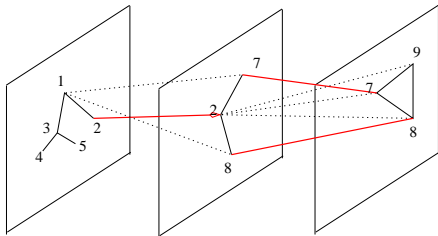
Finding communities that gather direct co-authors and also authors that share common co-authors.



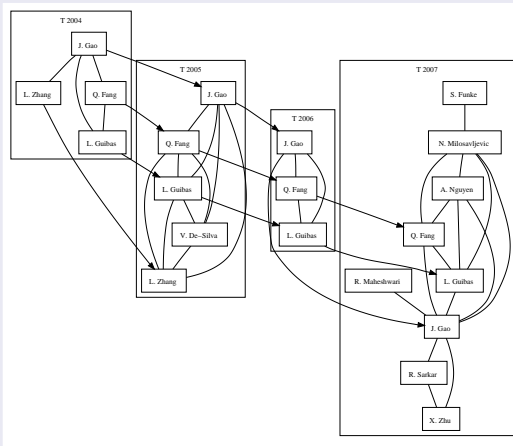
A single evolving graph

Dynamic communities

Finding communities that gather direct co-authors and also authors that share common co-authors.



Identification of evolving communities on Infocom data



Impact of the graph topology on the evolution of communities

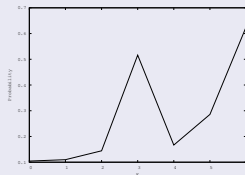
What is the impact of the graph topology on the evolution of a well identified community?

Infocom data

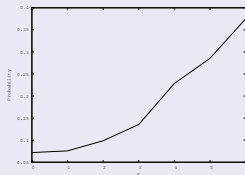
⇒ What are the dependencies between the co-authorship graph topology and being member of the program committee board?

- Dependencies on the number of co-authors that belong to the PC board
- Dependencies on a broader range of topology features

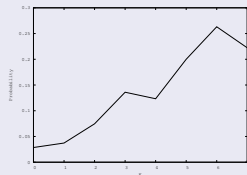
Are PC members co-opted by one of their relations/colleagues?



1995



2000



2005

Proportion $P(k)$ of authors who join Infocom PC as a function of co-authors who are already in the PC board.

Analyzing the dependencies between the co-author graph topology and the PC board community

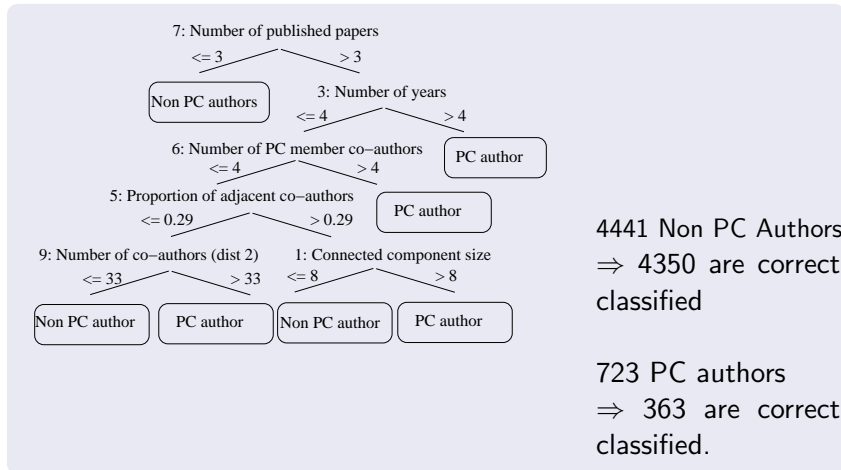
- Approach based on Backstrom et al. approach^a
- Features describing the graph topology are used to construct a predictive model of been member of the PC board
- Technique used : decision trees

^aL. Backstrom, D. Huttenlocher, J. Kleinberg and X. Lan, Group formation in large social networks: membership, growth, and evolution, KDD 2006.

Describing a co-authorship network

Feature number	Feature description
1	Connected component size
2	Number of co-authors (degree)
3	Number of years where an author has published at least one paper
4	Number of co-authors of a given author that are also co-authors together (similar to the clustering coefficient)
5	Proportion of co-authors of a given author that are also co-authors together
6	Number of co-authors that are PC members
7	Number of published papers
8	Number of co-authors at distance at most 2
9	Number of co-authors at distance at most 3

The obtained decision tree



Conclusion and future work

Summary

- Identification of stable communities based on frequent co-authorship structures
- Identification of evolving communities while taking into account temporal evolution
- Direct co-authors have a significant impact on the Infocom PC board composition

Future work

- Studying communities embedded in a larger network (arXiv, DBLP)
- Studying PC board of several conferences and determine if one can found a small world structure between different boards