Overlapping community detection in dynamic networks

Qinna WANG

LIP6

l'université Pierre et Marie Curie

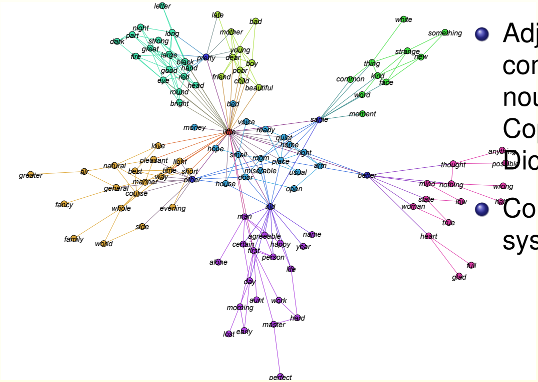
CNRS

12 Avril 2012

Contents

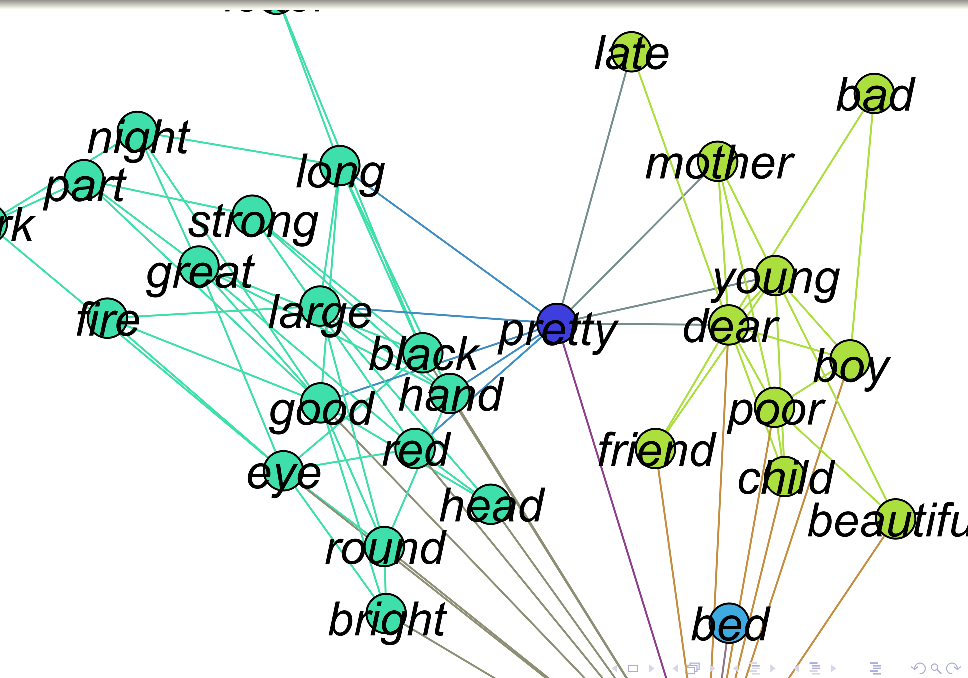
Community structure

A complex network has community structure if the nodes of the networks are easily grouped into sets of nodes such that each set of nodes is densely connected internally.



- Adjacency network of common adjectives and nouns in the novel David Copperfield by Charles Dickens [New06].
- Communities refer to system functions

Community structure



Community structure

Communities may overlap.

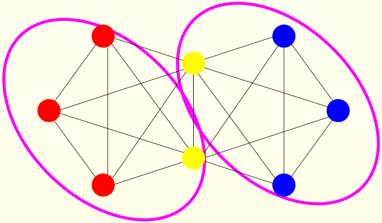
Overlapping community structure

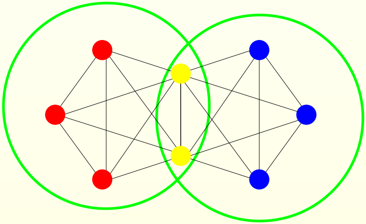
Partition

a division of a graph into disjoint communities, such that each node belongs to a unique community.

Cover

A division of a graph into overlapping (or fuzzy) communities, such that some nodes are shared by several communities.





Dynamic networks

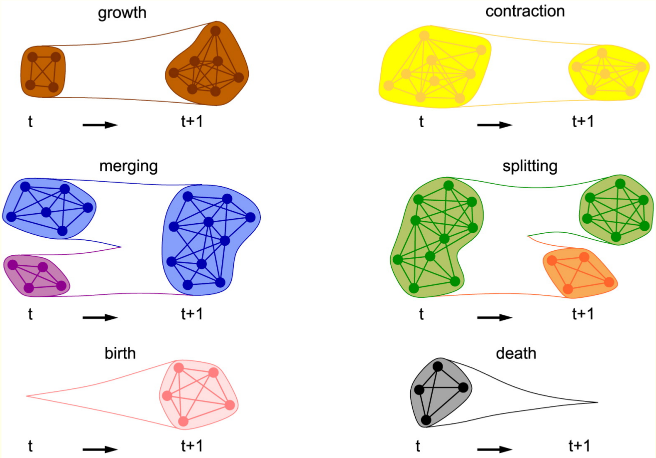
- Network structure dynamically evolves in time.
-
- Our second **problem** : Tracking community evolution in dynamic networks

Dynamic networks

- Network structure dynamically evolves in time.
-
- Our second **problem** : Tracking community evolution in dynamic networks

Community dynamics

Scenarios in the evolution of communities by *Gergely Palla et al.* [PBV07]



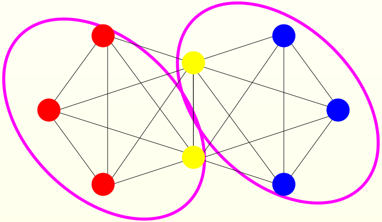
Our contributions

- Overlapping community detection
- Tracking community evolution and identifying community dynamics

Contents

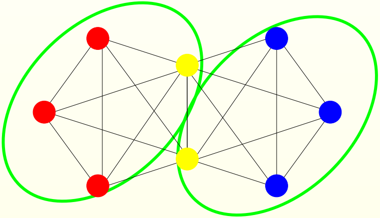
Fuzzy detection

- When running several times the Louvain algorithm on the same given network,



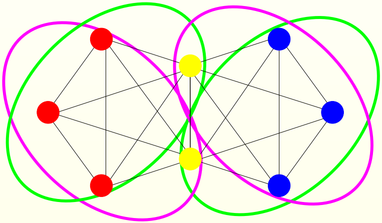
Fuzzy detection

- When running several times the Louvain algorithm on the same given network, we observe different partitions.



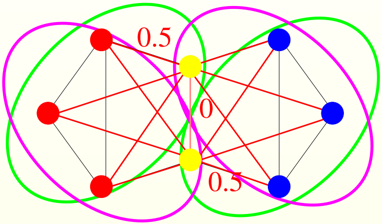
Fuzzy detection

- When running several times the Louvain algorithm on the same given network, we observe different partitions.
- "*Oscillating*" nodes are possible overlapping nodes.



Fuzzy detection

- When running several times the Louvain algorithm on the same given network, we observe different partitions.
- "*Oscillating*" nodes are possible overlapping nodes.
- Problem : compute $\mathbf{P}_c = [P_{c_i,c_j}]_{n\times k}$ whose P_{c_i,c_j} represents the probability of node c_i belonging to the community $\mathcal{C}_j$.



Fuzzy detection

Related work [1]

Definition

An edge $e = (i, j)$ is *external* if $p_{ij} < \alpha^*$(*e.g.* $\alpha^* = 99.5\%$)

Definition

A *robust cluster* is the composition of connected graph after removing external edges

Definition

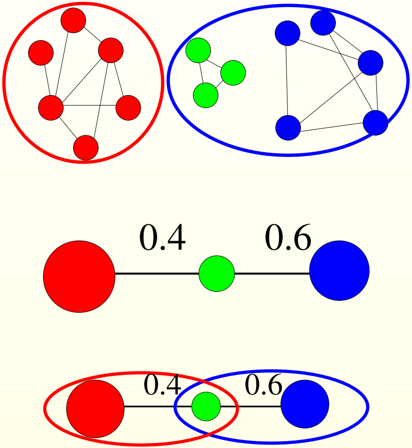
The core of the community is the robust cluster has the maximum number of nodes.

1. *Gfeller et al.*, Finding instabilities in the community structure of complex networks, Physical review. E

Fuzzy detection

1. Detect robust clusters
   - Several runs of Louvain algorithm to compute a co-appearance matrix $\mathbf{P} = [P_{ij}]_{n\times n}$
   - Save the partition $\mathcal{P}_{\text{opt}}$ with the highest modularity
   - Remove all external edges from $\mathcal{P}_{\text{opt}}$
2. Adjust memberships of robust clusters
   - Identify community core : $\widehat{c}_i = \arg\max_{c_j \subseteq \mathcal{C}_i} |c_j|$, where $\mathcal{C}_i \in \mathcal{P}_{\text{opt}}$
   - Compute $\mathbf{P}_c = [P_{c_i,c_j}]$
   - Add robust cluster c_i to community $\mathcal{C}_j$ if $p_{c_j,\widehat{c}_i} \geq \beta^*$(*e.g.* $\beta^* = 10\%$)

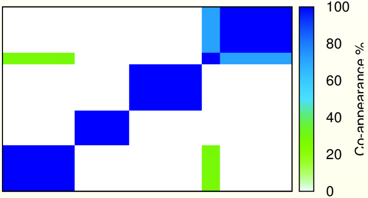
Fuzzy detection



Datasets

- Synthetic graphs containing hierarchical structure
    - 16 small groups : $k_1 = 30$
    - 4 super groups : $k_2 = 13$
        - 1 group (modular overlaps) is shared by 2 super groups
        - for others, each group belongs to a unique super group.
    - External links : $k_3 = 5$

Performances in testing artificial networks

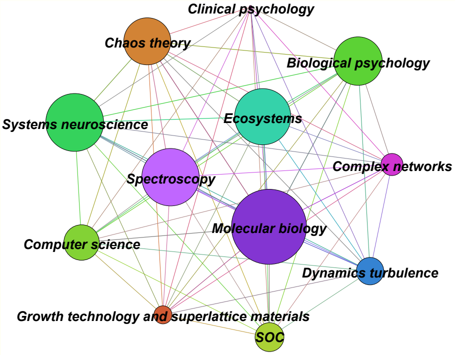


Conclusion : Fuzzy detection detects modular overlaps

Applications to Complex System Science

Complex System Science is a citation graph [2].

- Source : ISI Web of knowledge
- Node : an article (2000 - 2009) ; contains keywords (complex systems)
- Weight : bibliographic coupling [Kes63] :
$$w_{ij} = \frac{|R_i \cap R_j|}{\sqrt{|R_i|\,|R_j|}}.$$
- Communities : research topics or theoretical fields.

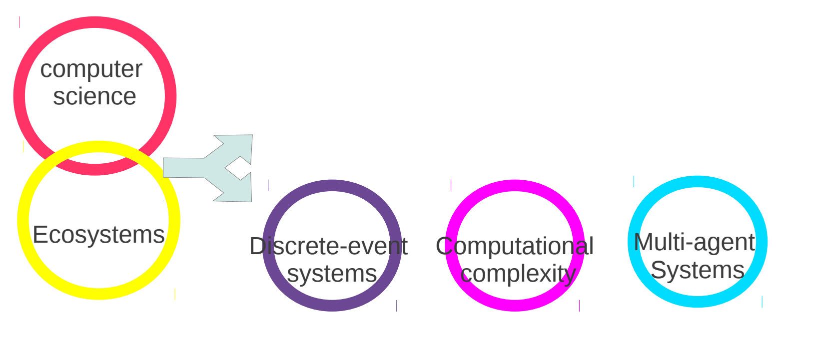


Results of Louvain algorithm

2. *Grauwin Sebastian, Jensen Pablo et al.* Complex Systems Science : Dreams of Universality, Reality of Interdisciplinarity. Journal of the American Society for Information

Applications to Complex System Science

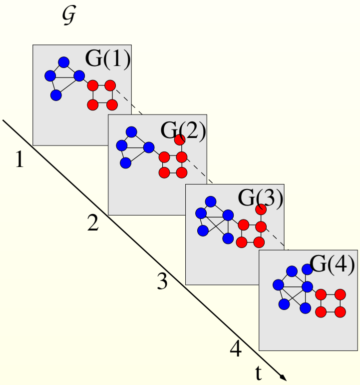
Results in views of modular overlaps



Contents

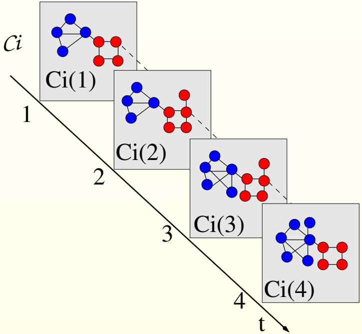
Introduction about community evolution

- A *dynamic graph* $\mathcal{G}(V, \mathcal{E})$ on a finite time sequence $1 \dots \Delta$ is a sequence of graph snapshots $\{G(1), \dots, G(\Delta)\}$.
- The evolution of a community can be tracked by its evolution path : $\mathrm{Evol}(\mathcal{C}_i) := \{C_i(\delta), \dots, C_i(\delta + \Delta)\}$

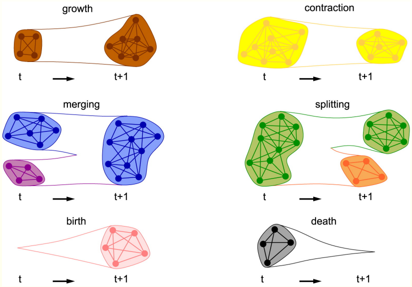


Introduction about community evolution

- A *dynamic graph* $\mathcal{G}(V, \mathcal{E})$ on a finite time sequence $1 \ldots \Delta$ is a sequence of graph snapshots $\{G(1), \ldots, G(\Delta)\}$.
- The evolution of a community can be tracked by its evolution path : $\mathrm{Evol}(\mathcal{C}_i) := \{C_i(\delta), \ldots, C_i(\delta + \Delta)\}$



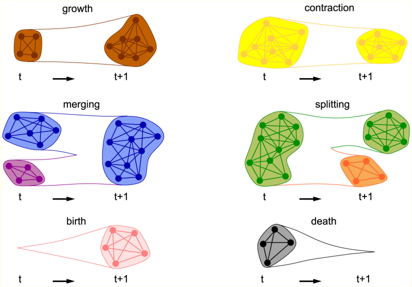
Community dynamics



Community dynamics make community evolution become difficult to track.

However, the definition of community dynamics is a problem.

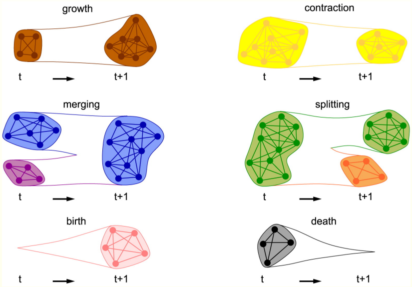
Community dynamics



Community dynamics make community evolution become difficult to track.

However, the definition of community dynamics is a problem.

Community dynamics



Community dynamics make community evolution become difficult to track.

However, the definition of community dynamics is a problem.

Group persistence two-stage method

1. Detect partitions, robust clusters by fuzzy detection algorithm at each time step ;
2. Map clusters through group persistence.

Our definition

- Given a temporal cluster $C_i(t)$ at time t, [3]

Definition (Community predecessor)

if $C_j(t-1)$ has the maximum overlap size at time $t-1$, such that $C_j(t-1) \rightarrow C_i(t)$

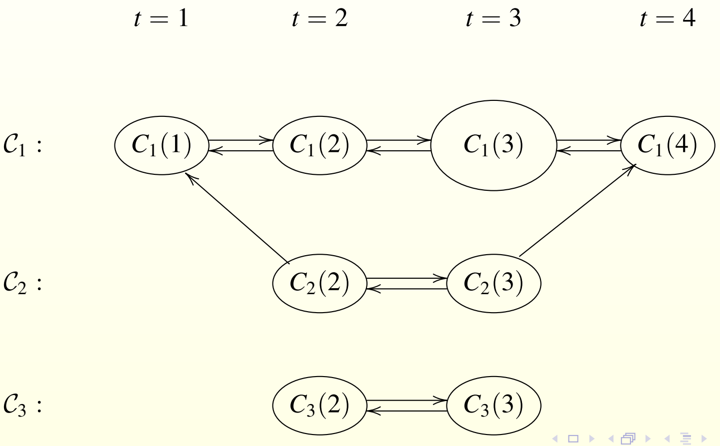
Definition (Community successor)

if $C_k(t+1)$ has the maximum overlap size at time $t+1$, such that $C_i(t) \leftarrow C_k(t+1)$

3. *Q.Wang and E.Fleury*, Understanding community evolution in Complex systems science, 1st International Workshop on Dynamicity, December 12, Collocated with OPODIS 2011, Toulouse, France

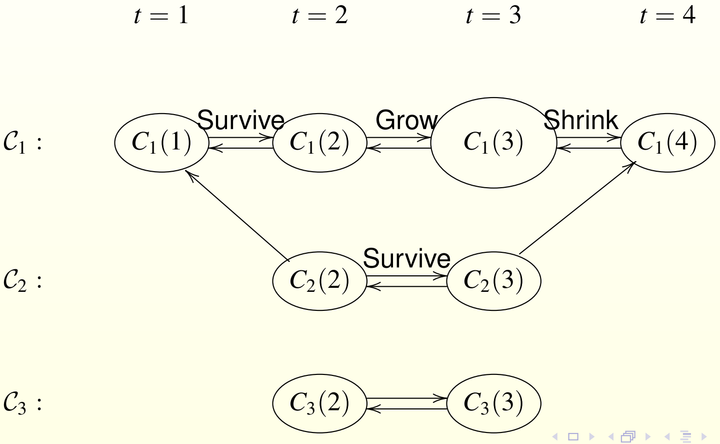
Asymmetrical relationship

- This asymmetrical property allows us to characterize community dynamics :



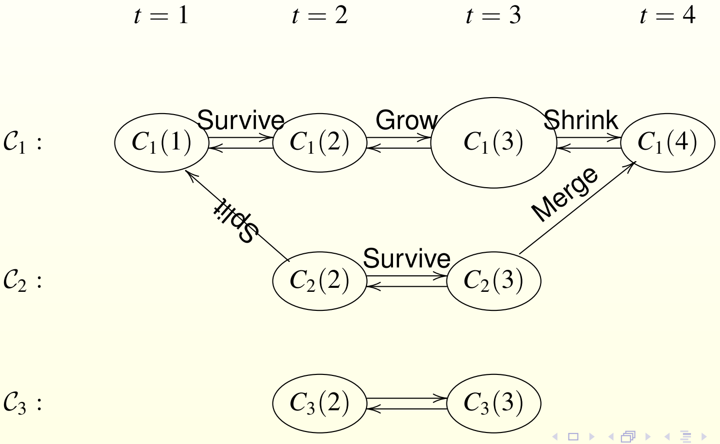
Asymmetrical relationship

- This asymmetrical property allows us to characterize community dynamics :



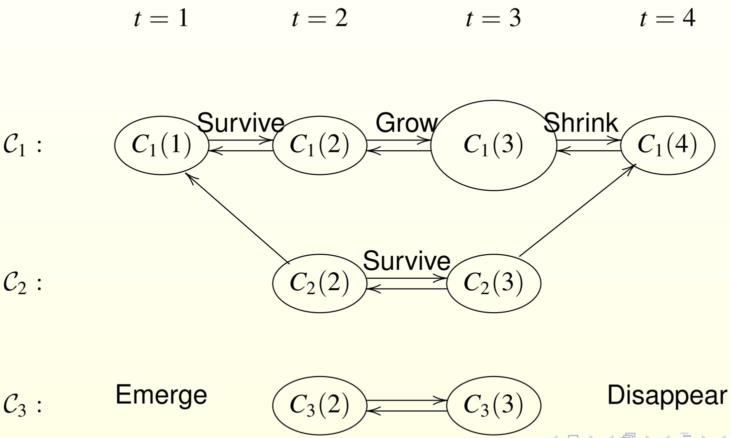
Asymmetrical relationship

- This asymmetrical property allows us to characterize community dynamics :

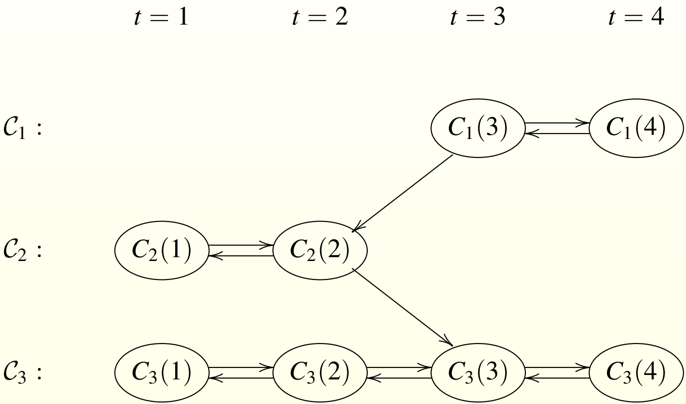


Asymmetrical relationship

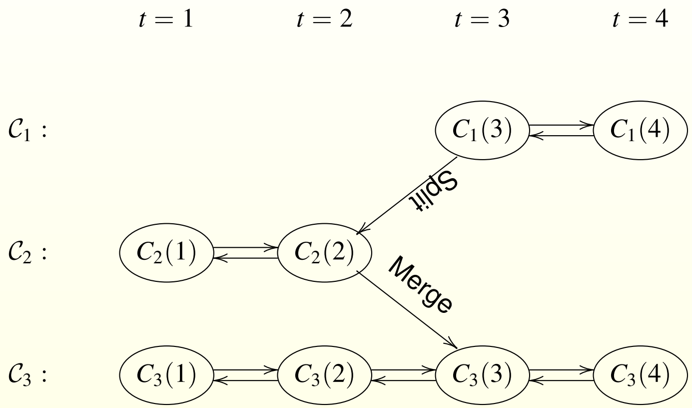
- This asymmetrical property allows us to characterize community dynamics :



Complex cases



Complex cases

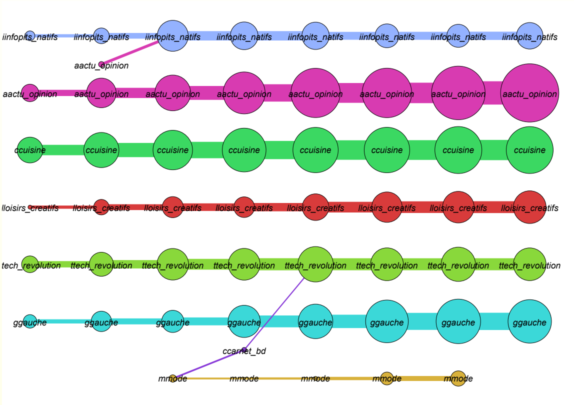


A dynamic blog network

- Dynamic blog networks [4] :
  - approximately six thousand blogs
  - Aggregate links between blogs every day(120 days)
  - 8 time steps (14 days as a time interval)

4. Abdelhamid Salah brahim, Matthieu Latapy *et al. Citations among blogs in a hierarchy of communities : method and case study*. Journal of Computational Science, Vol 2(3), 2011

Application to a dynamic blog network



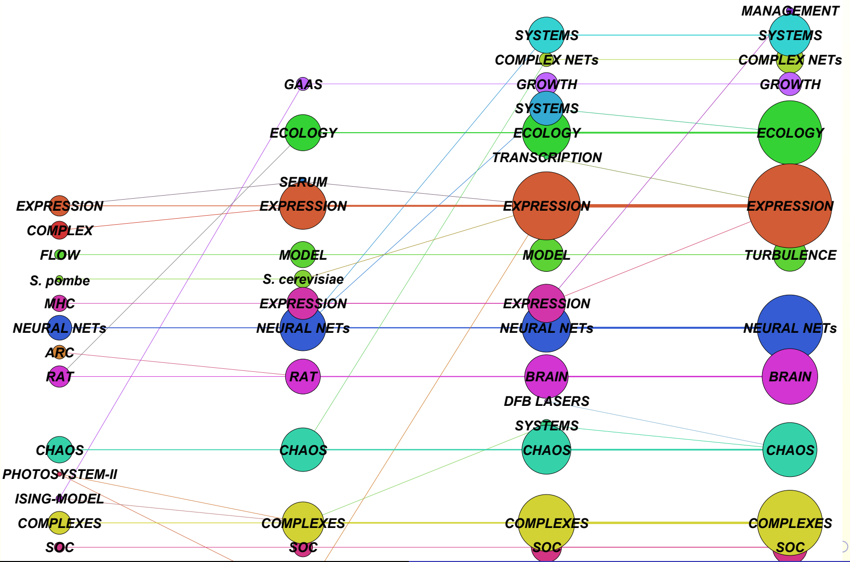
Each community whose evolution is survival shares the same topic.
New community corresponds to the event of new blogs.

The past history of complex system science network

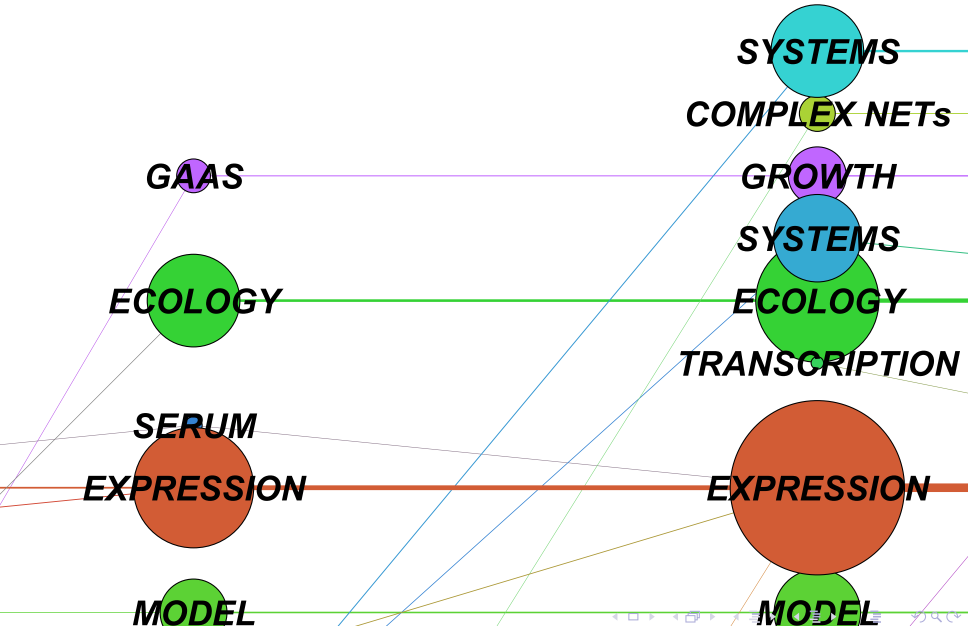
Time period	Number of nodes	Number of edges	Total weight
1985-1994	20286	1004458	183594
1990-1999	62040	6179802	1.0569e+06
1995-2004	109458	12662556	2.1206e+06
2000-2009	141163	19603888	3.6701e+06

TABLE: Properties of the past history of Complex System Sciences.

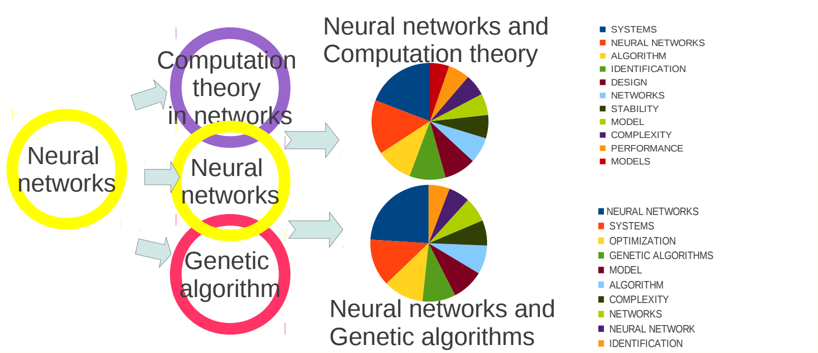
Application to a dynamic citation network



New scientific topics or fields



Split events in the past history of Complex System Sciences



Observation : The overlaps shared by split communities reveal their predecessor.

Contents

Conclusion and perspectives

Conclusion : We have explored computational techniques to study community organization of complex networks with overlapping nodes.

Future work :

- Visualization tool for overlapping community evolution.
- Add more constraints to smooth the shifts of community members.
- Analyse more dynamic networks : benchmarks for evaluating algorithms and structural properties in dynamic views.

List of publications

- International Conferences
  - Q.Wang and E.Fleury, *Detecting overlapping communities in graphs*, European conference on Complex Systems 2009, University of Warwick, UK
  - Q.Wang and E.Fleury, *Uncovering Overlapping Community Structure*, 2nd Workshop on Complex Networks, Brazil, 2010
  - Q.Wang and E.Fleury, *Mining time-dependent communities*, Latin-American Workshop on Dynamic Networks, Argentina, 2010
  - Q.Wang and E.Fleury, *Community detection with fuzzy community structure*, The First Workshop on Social Network Analysis in Applications, ASONAM 2011 :International Conference on Advances in Social Networks Analysis and Mining,Taiwan, 2011 (Best paper award)
  - Q.Wang and E.Fleury, *Understanding community evolution in Complex systems science*, 1st International Workshop on Dynamicity, December 12, Collocated with OPODIS 2011, Toulouse, France

Bibliographie

S. Asur, S. Parthasarathy, and D. Ucar.
An event-based framework for characterizing the evolutionary behavior of interaction graphs.
In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, page 921. ACM, 2007.

Kenneth P Burnham and David R Anderson.
Model selection and multimodel inference : a practical information-theoretic approach, volume 172.
Springer, 2002.

Albert-Laszlo Barabasi and Eric Bonabeau.
Scale-free networks.
Sci Am, 288(5) :60–69, May 2003.

Vladimir Batagelj, Patrick Doreian, and Anuška Ferligoj.
Generalized blockmodeling of two-mode network data.
Social Networks, 26(1) :29–53, 2004.

V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre.
Fast unfolding of communities in large networks.
Journal of Statistical Mechanics-Theory and Experiment, 2008.

Michelle Girvan and M. E. J. Newman.
Community structure in social and biological networks.
Proc. Natl. Acad. Sci. USA, 99, :7821–7826, 2002.

Steve Gregory.
Finding overlapping communities in networks by label propagation.
New Journal of Physics, 12(10) :103018, 2010.

Manel Ben Jdidia, Céline Robardet, and Eric Fleury.
Communities detection and analysis of their dynamics in collaborative networks.
In *ICDIM*, pages 744–749. IEEE, 2007.

M M Kessler.
Bibliographic coupling between scientific papers.

American Documentation, 14(1) :10–25, 1963.

A. Lancichinetti, S. Fortunato, and J. Kertesz.
Detecting the overlapping and hierarchical community structure in complex networks.
New Journal of Physics, 11, 2009.

R. D. LUCE and A. D. PERRY.
A method of matrix analysis of group structure.
Psychometrika, 14(2) :95–116, Jun 1949.

Andrea Lancichinetti and Jose J Radicchi, Filippo Ramasco.
Statistical significance of communities in networks.
81, :046110, 2009.

Andrea Lancichinetti, Filippo Radicchi, and Santo Ramasco, Jose' Javier Fortunato.
Finding statistically significant communities in networks.
2010.

PJ Mucha, Thomas Richardson, Kevin Macon, and Mason A. Porter.
Community structure in time-dependent, multiscale, and multiplex networks.
science, 876 :10–13, 2010.

M. E. J. Newman.
Finding community structure in networks using the eigenvectors of matrices.
Phys. Rev. E, 74 :036104, Sep 2006.

Gergely Palla, Albert-Laszlo Barabasi, and Tamas Vicsek.
Quantifying social group evolution.
Nature, 446 :664–667, 2007.

Gergely Palla, Imre Derenyi, and Tamas Farkas, Illes Vicsek.
Uncovering the overlapping community structure of complex networks in nature and society.
435, :814, 2005.

Shuye Pu, Jessica Wong, Brian Turner, Emerson Cho, and Shoshana J Wodak.
Up-to-date catalogues of yeast protein complexes.
Nucleic Acids Res, 37(3) :825–831, Feb 2009.

Georg Simmel.
The persistence of social groups.
American Journal of Sociology, 3 (1897) : 662-698.

Lei Tang, Huan Liu, Jianping Zhang, and Zohreh Nazeri.
Community evolution in dynamic multi-mode networks.
In *International Conference on Knowledge Discovery and Data Mining*, page 8, 2008.

D. J. Watts and S. H. Strogatz.
Collective dynamics of 'small-world' networks.
Nature, 393(6684) :440–442, Jun 1998.

Tianbao Yang, Yun Chi, Shenghuo Zhu, Yihong Gong, and Rong Jin.

A bayesian approach toward finding communities and their evolutions in dynamic social networks.
In *SIAM Conference on Data Mining (SDM)*, 2009.