Overlapping community detection in dynamic networks

Qinna WANG

LIP6

l'université Pierre et Marie Curie
CNRS

12 Avril 2012
Contents

1. Background
2. Overlapping community detection
3. Community evolution in dynamic networks
4. Conclusion and perspectives
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

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.
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]

- **Growth**
- **Contraction**
- **Merging**
- **Splitting**
- **Birth**
- **Death**
Our contributions

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

1. Background
2. Overlapping community detection
3. Community evolution in dynamic networks
4. Conclusion and perspectives
Fuzzy detection

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

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

"Oscillating" nodes are possible overlapping nodes.

Problem: compute

\[ P_c = [P_{c_i,c_j}]_{n \times k} \]

whose \( P_{c_i,c_j} \) represents the probability of node \( c_i \) belonging to the community \( C_j \).
Fuzzy detection

Related work

**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.

---

Fuzzy detection

1. Detect robust clusters
   - Several runs of Louvain algorithm to compute a co-appearance matrix $P = [P_{ij}]_{n \times n}$
   - Save the partition $P_{opt}$ with the highest modularity
   - Remove all external edges from $P_{opt}$

2. Adjust memberships of robust clusters
   - Identify community core: $\hat{c}_i = \arg \max_{c_j \subseteq C_i} |c_j|$, where $C_i \in P_{opt}$
   - Compute $P_c = [P_{c_i,c_j}]$
   - Add robust cluster $c_i$ to community $C_j$ if $p_{c_j,\hat{c}_i} \geq \beta^*(e.g. \beta^* = 10\%)$
Fuzzy detection

Overlapping community detection
Community evolution in dynamic networks
Conclusion and perspectives

Fuzzy detection
Applications to real networks

Qinna WANG

Overlapping community detection in dynamic networks
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: $k_3 = 5$
Conclusion: Fuzzy detection detects modular overlaps
Applications to Complex System Science

Complex System Science is a citation graph.

- **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**

Results in views of modular overlaps
A dynamic graph $\mathcal{G}(V, E)$ 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: $\text{Evol}(C_i) := \{C_i(\delta), \ldots, C_i(\delta + \Delta)\}$.
Introduction about community evolution

- A *dynamic graph* $\mathcal{G}(V, E)$ 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:
  \[\text{Evol}(C_i) := \{C_i(\delta), \ldots, C_i(\delta + \Delta)\}\]
Community dynamics make community evolution become difficult to track.

However, the definition of community dynamics is a problem.
Community dynamics make community evolution become difficult to track.

However, the definition of community dynamics is a problem.
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.
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:

\[ t = 1 \quad t = 2 \quad t = 3 \quad t = 4 \]

\[ C_1 : \]

\[ C_1(1) \rightarrow C_1(2) \rightarrow C_1(3) \rightarrow C_1(4) \]

\[ C_2 : \]

\[ C_2(2) \rightarrow C_2(3) \]

\[ C_3 : \]

\[ C_3(2) \rightarrow C_3(3) \]
This **asymmetrical** property allows us to characterize community dynamics:

\[ t = 1 \quad t = 2 \quad t = 3 \quad t = 4 \]

- \( C_1 \):
  - \( C_1(1) \) \( \xrightarrow{\text{Survive}} \) \( C_1(2) \) \( \xrightarrow{\text{Grow}} \) \( C_1(3) \) \( \xrightarrow{\text{Shrink}} \) \( C_1(4) \)

- \( C_2 \):
  - \( C_2(2) \) \( \xrightarrow{\text{Survive}} \) \( C_2(3) \)

- \( C_3 \):
  - \( C_3(2) \) \( \xrightarrow{\text{}} \) \( C_3(3) \)
This asymmetrical property allows us to characterize community dynamics:

\[ t = 1 \quad \quad \quad t = 2 \quad \quad \quad t = 3 \quad \quad \quad t = 4 \]

**C_1:**
- \( C_1(1) \) Survive to \( C_1(2) \) Grow to \( C_1(3) \) Shrink to \( C_1(4) \)
- \( C_1(1) \) Split to \( C_2(2) \) Survive to \( C_2(3) \)

**C_2:**
- \( C_2(2) \) Survive to \( C_2(3) \)

**C_3:**
- \( C_3(2) \) to \( C_3(3) \)
Asymmetrical relationship

- This **asymmetrical** property allows us to characterize community dynamics:

\[
\begin{align*}
  t & = 1 & t & = 2 & t & = 3 & t & = 4 \\
  C_1 : \quad & \text{Survive} & \quad & \text{Grow} & \quad & \text{Shrink} & \\
  C_1(1) & \rightarrow & C_1(2) & \rightarrow & C_1(3) & \rightarrow & C_1(4) \\
  C_2 : \quad & \text{Survive} & \\
  C_2(2) & \rightarrow & C_2(3) \\
  C_3 : \quad & \text{Emerge} & \quad & \text{Disappear} & \\
  C_3(2) & \rightarrow & C_3(3)
\end{align*}
\]
Complex cases

\[ t = 1 \quad t = 2 \quad t = 3 \quad t = 4 \]

\[ C_1: \]
\[ C_2: \]
\[ C_3: \]
Complex cases

$t = 1$ \hspace{1cm} $t = 2$ \hspace{1cm} $t = 3$ \hspace{1cm} $t = 4$

$C_1 :$

$C_2 :$

$C_3 :$
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)

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

<table>
<thead>
<tr>
<th>Time period</th>
<th>Number of nodes</th>
<th>Number of edges</th>
<th>Total weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985-1994</td>
<td>20286</td>
<td>1004458</td>
<td>183594</td>
</tr>
<tr>
<td>1990-1999</td>
<td>62040</td>
<td>6179802</td>
<td>1.0569e+06</td>
</tr>
<tr>
<td>1995-2004</td>
<td>109458</td>
<td>12662556</td>
<td>2.1206e+06</td>
</tr>
<tr>
<td>2000-2009</td>
<td>141163</td>
<td>19603888</td>
<td>3.6701e+06</td>
</tr>
</tbody>
</table>

**Table:** Properties of the past history of Complex System Sciences.
Application to a dynamic citation network
New scientific topics or fields

- SYSTEMS
- SYSTEMS
- COMPLEX NETs
- GROWTH
- TRANScription
- TRANSCRIPTION
- ECOLOGY
- ECOLOGY
- SERUM
- EXPRESSION
- MODEL
- GAAS
- EXPRESSION
- EXPRESSION
- MODEL
- MODEL
- SYSTEMS
Split events in the past history of Complex System Sciences

Observation: The overlaps shared by split communities reveal their predecessor.
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, *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
S. Asur, S. Parthasarathy, and D. Ucar.
An event-based framework for characterizing the evolutionary behavior of interaction graphs.

Kenneth P Burnham and David R Anderson.

Albert-Laszlo Barabasi and Eric Bonabeau.
Scale-free networks.

Vladimir Batagelj, Patrick Doreian, and Anuška Ferligoj.
Generalized blockmodeling of two-mode network data.


M M Kessler. Bibliographic coupling between scientific papers.


PJ Mucha, Thomas Richardson, Kevin Macon, and Mason A. Porter.
Community structure in time-dependent, multiscale, and multiplex networks.

Finding community structure in networks using the eigenvectors of matrices.

Quantifying social group evolution.

Gergely Palla, Imre Derenyi, and Tamas Farkas, Illes Vicsek.
Uncovering the overlapping community structure of complex networks in nature and society.
Shuye Pu, Jessica Wong, Brian Turner, Emerson Cho, and Shoshana J Wodak.
Up-to-date catalogues of yeast protein complexes.  

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.  

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.*