Multiplex networks analysis

Rushed Kanawati

$A^3$, LIPN, CNRS UMR 7030
SPC - University Paris Nord

http://lipn.fr/~kanawati

rushed.kanawati@lipn.univ-paris13.fr

FRUMAM, Marseilles
4 February 2015

source: muxviz
INTRODUCTION

Multiplex network analysis

Community Detection in Multiplex Networks
  - Adaptation of monoplex community detection algorithms
  - MuxLicod Algorithm
  - Experiments

Link prediction in multiplex networks

Conclusion
Different *interaction* networks with similar topological features: *sparsity, small world, power-law degree distribution, high clustering coefficient, community structure, etc.*
A lot of interesting results using a simple model: interacting nodes
- Node’s characterization: centralities
- Diffusion models on simple networks
- Community detection
- Link prediction models
- …

Network science is mature enough to move towards more complex models
**MULTIPLEX NETWORK**

**Definition**
A set of nodes related by different types of relations

**Motivation**
- Real networks are **dynamic**.
- Real networks are **heterogeneous**.
- Nodes are usually **qualified** by a set of attributes.

Source: muxviz
MULTIPLEX NETWORKS: RELATED TERMS

Recommended readings

POWER OF MULTIPLEX MODEL

Multi-relationnal networks

European airports network
POWER OF MULTIPLEX MODEL

Dynamic networks

Academic collaborations per year
POWER OF MULTIPLEX MODEL

Attributed networks

Teenage friendship network - Behavioral attributes: Sport practice level, Alcohol, Tobacco & Cannabis consumption

Similarity graphs can be defined over nodes using attribute-similarity measures: $\beta$-threshold graphs, $knn$-graphs, Relative neighborhood graphs
Heterogeneous networks

DBLP author-centred multiplex network
**Multiplex Network: Notations**

\[ G = \langle V, E_1, \ldots, E_\alpha : E_k \subseteq V \times V \ \forall k \in \{1, \ldots, \alpha\} \rangle \]

- **\( V \):** set of nodes (a.k.a. vertices, actors, sites)
- **\( E_k \):** set of edges of type \( k \) (a.k.a. ties, links, bonds)

**Notations**

- **\( A^{[k]} \):** Adjacency Matrix of slice \( k \) : \( a^{[k]}_{ij} \neq 0 \) if the nodes \((v_i, v_j) \in E_k\), 0 otherwise.
- **\( m^{[k]} = |E_k| \).** We have often \( m \sim n \)
- **Neighbor’s of \( v \) in slice \( k \):** \( \Gamma(v)^{[k]} = \{ x \in V : (x, v) \in E_k \} \).
- **All neighbors of \( v \):** \( \Gamma(v)^{tot} = \bigcup_{s \in \{1, \ldots, \alpha\}} \Gamma(v)^{[s]} \)
- **Node degree in slice \( k \):** \( d^k_v = \| \Gamma(v)^{[k]} \| \)
- **Total degree of node \( v \):** \( d^{tot}_v = \| \Gamma^{tot}(v) \| \)
MULTIPLEX NETWORK ANALYSIS

1. Node-related tasks
   Degree, centralities, neighborhood, dyadic metrics

2. Community-related tasks
   Graph partitioning, overlapping communities, local communities

3. Network-related tasks
   Link prediction
**Analysis approaches**

1. **Transformation into a monoplex centred problem**
   - Layer aggregation approaches.
   - Hypergraph transformation based approaches
   - *Ensemble approaches*

2. **Generalization of monoplex oriented algorithms to multiplex networks.**
Layer Aggregation
Layer Aggregation

Aggregation functions

\[ A_{ij} = \begin{cases} 1 & \exists 1 \leq l \leq \alpha : A_{ij}^{[l]} \neq 0 \\ 0 & \text{otherwise} \end{cases} \]

\[ A_{ij} = \| \{ d : A_{ij}^{[d]} \neq 0 \} \| \]

\[ A_{ij} = \frac{1}{\alpha} \sum_{k=1}^{\alpha} w_k A_{ij}^{[k]} \]

\[ A_{ij} = \text{sim}(v_i, v_j) \]
K-uniform hypergraph transformation

Principle

- A k-uniform hypergraph is a hypergraph in which the cardinality of each hyperedge is exactly \( k \).

- Mapping a multiplex to a 3-uniform hypergraph \( \mathcal{H} = (\mathcal{V}, \mathcal{E}) \) such that:

\[
\mathcal{V} = V \cup \{1, \ldots, \alpha\}
\]

\[
(u, v, i) \in \mathcal{E} \text{ if } \exists l : A_{uv}^{[l]} \neq 0, u, v \in V, i \in \{1, \ldots, \alpha\}
\]

- Apply hypergraphs analysis approaches (Ex. tensor-based approaches)
**MULTIPLEX: NODE DEGREE**

Some options

- \[ d_{v}^{\text{mux}} = \frac{1}{\alpha} \sum_{k=1}^{\alpha} d_{v}^{[k]} \]
- \[ d_{v}^{\text{mux}} = - \sum_{k=1}^{\alpha} \frac{d_{v}^{[k]}}{d_{v}^{[\text{tot}]} log \left( \frac{d_{v}^{[k]}}{d_{v}^{[\text{tot}]}} \right)} \]
- \[ \ldots \]
- \[ d_{v}^{\text{mux}} = |\Gamma^{\text{mux}}(v)| \]
**Multiplex: Node neighborhood**

Some options

- $\Gamma^{mux}(v) = \bigcup_{k=1}^{\alpha} \Gamma^k(v)$
- $\Gamma^{mux}(v) = \bigcap_{k=1}^{\alpha} \Gamma^k(v)$
- $\Gamma^{mux}(v) = \left\{ x \in \Gamma(v)^{tot} : \text{sim}(x, v) \geq \delta \right\} \delta \in [0, 1]$
- $\Gamma^{mux}(v) = \left\{ x \in \Gamma(v)^{tot} : \frac{\Gamma(v)^{tot} \cap \Gamma(x)^{tot}}{\Gamma(v)^{tot} \cup \Gamma(x)^{tot}} \geq \delta \right\}$
- ...
MULTIPLEX: DYADIC MEASURES

Some options

- $X_{average} = \frac{\sum_{\alpha=1}^{m} X(u,v)^{[\alpha]} }{m} \quad \forall u,v \in V$ and $(u,v) \notin E_i.$

- $X_{ent}(u,v) = - \sum_{\alpha=1}^{m} \frac{X(u,v)^{[\alpha]} }{X_{total}} \log(\frac{X(u,v)^{[\alpha]} }{X_{total}})$ where $X_{total} = \sum_{\alpha=1}^{m} X(u,v)^{[\alpha]}.$

- $\ldots$
**COMMUNITY?**

Some definitions:

- A dense subgraph loosely coupled to other modules in the network
- A community is a set of nodes seen as one by nodes outside the community
- A subgraph where almost all nodes are linked to other nodes in the community.

What is a dense subgraph in a multiplex network?

BerlingerioCG11
Community detection in multiplex networks

Approaches

1. Transformation into a monoplex community detection problem
   - Layer aggregation approaches.
   - Multi-objective optimization approach.
   - Ensemble clustering approaches

2. Generalization of monoplex oriented algorithms to multiplex networks.
   - Generalized-modularity optimization
   - Seed-centric approaches
MULTI-OBJECTIVE OPTIMIZATION APPROACH

1. Rank the set of $\alpha$ layers according to some importance criteria
2. $C_1 \leftarrow \text{community}(G^{[1]})$
3. for $i \in [2, \alpha]$ do:
   $C_i \leftarrow \text{optimize} (\text{community}(G^{[i]}), \text{similarity}(C_{i-1}))$
4. return $C_\alpha$
ENSEMBLE CLUSTERING APPROACHES
ENSEMBLE CLUSTERING APPROACHES

- CSPA: Cluster-based Similarity Partitioning Algorithm
- HGPA: HyperGraph-Partitioning Algorithm
- MCLA: Meta-Clustering Algorithm
- ...
ENSEMBLE CLUSTERING: APPROACHES

CSPA: Cluster-based Similarity Partitioning Algorithm

- Let \( K \) be the number of basic models, \( C_i(x) \) be the cluster in model \( i \) to which \( x \) belongs.

- Define a similarity graph on objects: \( \text{sim}(v, u) = \frac{\sum_{i=1}^{K} \delta(C_i(v), C_i(u))}{K} \)

- Cluster the obtained graph:
  - Isolate connected components after pruning edges
  - Apply community detection approach

- Complexity: \( \mathcal{O}(n^2kr) \): \( n \) # objects, \( k \) # of clusters, \( r \)# of clustering solutions
CSPA : Exemple

from Seifi, M. Cœurs stables de communautés dans les graphes de terrain. Thèse de l’université Paris 6, 2012
ENSEMBLE CLUSTERING: APPROACHES

HGPA: HyperGraph-Partitioning Algorithm

- Construct a hypergraph where nodes are objects and hyperedges are clusters.
- Partition the hypergraph by minimizing the number of cut hyperedges.
- Each component forms a meta cluster.
- Complexity: $O(nkr)$
ENSEMBLE CLUSTERING: APPROACHES

MCLA: Meta-Clustering Algorithm
- Each cluster from a base model is an item
- Similarity is defined as the percentage of shared common objects
- Conduct meta-clustering on these clusters
- Assign an object to its most associated meta-cluster
- Complexity: $O(nk^2r^2)$
Generalized modularity

\[
Q_{\text{multiplex}}(P) = \frac{1}{2\mu} \sum_{c \in P} \sum_{i,j \in c} \left( A_{ij}^{[s]} - \lambda_k \frac{d_i^{[k]} d_j^{[k]}}{2m^{[k]}} \right) \delta_{kl} + \delta_{ij} C_{ij}^{kl}
\]

\[
\mu = \sum_{j \in V} m^{[k]} + C_{jk}^l
\]

\[
C_{ij}^{kl} \text{ Inter slice coupling } = 0 \forall i \neq j
\]
MODULARITY OPTIMIZATION LIMITATIONS

Hypothesis

- The best partition of a graph is the one that maximize the modularity.
- If a network has a community structure, then it is possible to find a precise partition with maximal modularity.
- If a network has a community structure, then partitions having high modularity values are structurally similar.

All three hypothesis do not hold Good10, LAN11a.
SEED-CENTRIC ALGORITHMS

Algorithm 1: General seed-centric community detection algorithm

Require: $G = (V, E)$ a connected graph,

1: $C \leftarrow \emptyset$
2: $S \leftarrow \text{compute_seeds}(G)$
3: for $s \in S$ do
4: \hspace{1em} $C_s \leftarrow \text{compute_local_com}(s,G)$
5: \hspace{1em} $C \leftarrow C + C_s$
6: end for
7: return $\text{compute_community}(C)$
THE LICOD ALGORITHM [YK14]

1. Compute a set of seeds that are likely to be leaders in their communities
   
   *Heuristic*: nodes having higher degree centralities than their neighbors

2. Each node in the graph ranks seeds in function of its own preference
   
   *In function of increasing Shortest path*

3. Iterate till convergence: Each node modifies its preference vector in function of neighbor’s preferences
   
   *Applying rank aggregation methods.*
**MuxLicod**

**Multiplex degree centrality**

\[
d_i^{\text{multiplex}} = - \sum_{k=1}^{\alpha} \frac{d_i^{[k]}}{d_i^{[\text{tot}]}} \log \left( \frac{d_i^{[k]}}{d_i^{[\text{tot}]}} \right)
\]

**Multiplex shortest path**

\[
SP(u, v)^{\text{multiplex}} = \frac{\sum_{k=1}^{\alpha} SP(u, v)^{[k]}}{\alpha}
\]

**Multiplex neighborhood**

\[
\Gamma^{\text{mux}}(v) = \{ x \in \Gamma(v)^{\text{tot}} : \frac{\Gamma(v)^{\text{tot}} \cap \Gamma(x)^{\text{tot}}}{\Gamma(v)^{\text{tot}} \cup \Gamma(x)^{\text{tot}}} \geq \delta \}
\]
DATASETS

benchmark networks
Lazzega Lawyer network
#nodes 71
#layer 3
Datasets

Dataset
Physicians collaboration network
#nodes 246
#layers 3
Dataset

Bibsonomy
#nodes 361
#layers 2
DATASETS

Dataset

Vicker’s network [VC81]

#nodes 29
#layers 3
EVALUATION CRITERIA

1. Multiplex modularity
2. Redundancy [BCG11]

\[ \rho(c) = \sum_{(u,v) \in \bar{P}_c} \frac{\| \{k : \exists A_{uv}^{[k]} \neq 0 \} \|}{\alpha \times \| P_c \|} \]

\( \bar{P} \) the set of couple \((u, v)\) which are directly connected in at least two layers
RESULTS: REDUNDANCY
RESULTS: MULTIPLEX MODULARITY
Comparison with GenLouvain

Experiments on DBLP dataset.

Redundancy results

Multiplex modularity results
MULTIPLEX NETWORKS: SUPERVISED DYADIC LINK PREDICTION APPROACH

\[ G_{t_0} \quad G_{t_1} \quad G_{t_2} \quad \ldots \ldots \quad G_{t_k} \quad \ldots \ldots \quad G_{t_n} \]

\[ t_0 \quad t_1 \quad t_2 \quad \ldots \ldots \quad t_k \quad \ldots \ldots \quad t_n \]

Learning \quad Labeling

Training

Testing

\[ G_{learn} = \bigcup_{t=t_0}^{t_{k-2}} G_t \]

\[ G_{label} = \bigcup_{t=t_{k-1}}^{t_k} G_t \]
EXPERIMENTS: DBLP

<table>
<thead>
<tr>
<th>Years</th>
<th>Properties</th>
<th>Co-Author</th>
<th>Co-Venue</th>
<th>Co-Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970-1973</td>
<td>Nodes</td>
<td>91</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Edges</td>
<td>116</td>
<td>1256</td>
<td>171</td>
</tr>
<tr>
<td>1972-1975</td>
<td>Nodes</td>
<td>221</td>
<td>221</td>
<td>221</td>
</tr>
<tr>
<td></td>
<td>Edges</td>
<td>319</td>
<td>5098</td>
<td>706</td>
</tr>
<tr>
<td>1974-1977</td>
<td>Nodes</td>
<td>323</td>
<td>323</td>
<td>323</td>
</tr>
<tr>
<td></td>
<td>Edges</td>
<td>451</td>
<td>9831</td>
<td>993</td>
</tr>
</tbody>
</table>

Table: Basic statistics about the 3-layer DBLP multiplex networks

<table>
<thead>
<tr>
<th>Years</th>
<th># Positive</th>
<th># Negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train/Test</td>
<td>Labeling</td>
<td>16</td>
</tr>
</tbody>
</table>

Table: # examples extracted from co-authorship layer (number of unconnected nodes in connected components)
## Link prediction: Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-measure</td>
<td>AUC</td>
</tr>
<tr>
<td>$Set_{direct}$</td>
<td>0.0357</td>
<td>0.5263</td>
</tr>
<tr>
<td>$Set_{direct+indirect}$</td>
<td>0.0256</td>
<td>0.5372</td>
</tr>
<tr>
<td>$Set_{direct+multiplex}$</td>
<td>0.0592</td>
<td>0.5374</td>
</tr>
<tr>
<td>$Set_{all}$</td>
<td>0.0153</td>
<td>0.5361</td>
</tr>
<tr>
<td>$Set_{multiplex}$</td>
<td>0.0374</td>
<td>0.5181</td>
</tr>
</tbody>
</table>

Table: Comparative link prediction results applying decision tree algorithm using different types of attributes
CONCLUSIONS

- **Multiplex networks** provide a rich representation of real-world interaction systems
- **Promising community detection approaches**: Local approaches (seed-centric)
- **A lot of work to reformulate basic network concepts for multiplex settings.**
- **Problems**: Evaluation and interpretation of computed communities: Recommendation-task based evaluation!
- **Ideas to explore:**
  - Multiplex approach for enhancing community detection in monoplex networks
  - Ensemble selection approaches
  - Graph coarsening
  - Multiplex of multiplexes
Research in modeling, analyzing and mining large-scale networks has attracted an increasing effort in the last few years. A major trend of work in network modeling and mining concerns analyzing homogeneous static networks. However, in real world settings, networks are often dynamic, snapshot of a network. Multiplex network analysis.
PERSONAL RELATED BIBLIOGRAPHY

That’s all folks!

Questions?
BIBLIOGRAPHY I


