Multiplex Network Mining: A Brief Survey

Rushed Kanawati

***Abstract*—Multiplex network model has been recently proposed as a mean to capture high level complexity in real-world interaction networks. A multiplex network can roughly be defined as a multi-layer network. Each layer contains the same set of nodes but a different type of links. In spite of its simplicity, the model allows handling multi-relationnal, heterogeneous, dynamic and even attributed networks. However, working with multiplex networks requiers redefining and adapting almost all basic metrics and algorithms generally used to analyse complex networks. In this paper we provide an overview of recent algorithmic advances in mining and analyzing multiplex networks. We review also some publicly available tools for multiplex network analysis.**

***Index Terms*—Complex networks, Multiplex network, Multi relationnal networks, community detection, link prediction**

I. Introduction

NETWORKS have proved to be a useful tool to model structural complexity of a variety of complex systems in different domains including sociology, biology, ethology and computer science. Most studies until recently have focused on analyzing simple static networks. However, real complex network are heterogeneous (nodes and links may have different types) and/or dynamic. For example, in a social network, people are linked with different types of ties: friendship, family relationship, professional relationship,. . . , etc. Moreover these relationships may evolve with time. The concept of multiplex networks has been introduced with the goal to provide an expressive model for modeling real-world complex networks [1], [2], [3]. A multiplex network is roughly defined as a multi-layer graph where each layer contains the same set of nodes but interconnected by different types of links.

Figure 1 illustrates an exemple of a multiplex network. This is a 3-layer social network where layers represent *advice, friendship* and *co-work* relationships among partners and associates of a corporate law firm [4].

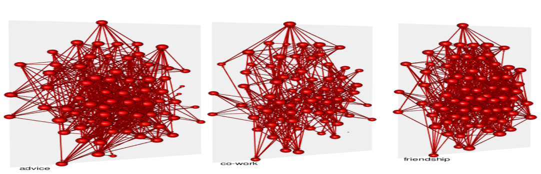

Fig. 1. Lazega law-firm network - visualization performed using muxviz package [5]

This simple extension of the basic graph model is powerful enough though to allow modeling different types of networks including:

R. Kanawati is an associate professor at LIPN CNRS UMR 7030, University Paris 13 e-mail: rushed.kanawati@lipn.fr

- *multi-relational network:* where each layer encodes one relation type,
- *dynamic network:* where a layer corresponds to the network state at a given time stamp,
- *attributed network:* where additional layers can be defined over the node set as a similarity graph induced by a similarity measure applied to the set of node's attributes.

However, analysis of multiplex networks requiers redefining most of the basic concepts and metrics usually used for complex network analysis including: node's degree, neighborhood, paths and node's centralities [6], [7], [3]. It requiers also providing new algorithms to handle basic complex networks analysis tasks such as community detection [8] and link prediction [9]. In this paper, we provide a brief review of recent algorithmic advances for multiplex network analysis and mining. In section II a formal definition of multiplex networks is provided and basic used notations are introduced. Section III defines basic node characterization metrics in multiplex networks. In the following section main algorithms for community detection in multiplex networks are reviewed. Section V gives brief informations about available multiplex network analysis tools. Finally we conclude in section VI

II. Definitions and notations

A multiplex network is defined as a triplet $G =< V, \mathbb{E}, C >$ where V is a set of nodes, $\mathbb{E} = \{E_1, \dots, E_\alpha\}$ is a set of α types of edges between nodes in V. We have $E_k = \{(v_i, v_j) : i \neq j, v_i, v_j \in V\}$. C is the set of coupling links that represent links between a node and itself across different layers. We have $C = \{(v, v, l, k) : v \in V, l, k \in [1, \alpha], l \neq k\}$. Where (v, v, l, k) denotes a link from node v in layer l to node v in layer k. Different coupling schemes can be applied. Figure 2 illustrates the two most basic couplings :

- *Ordinal coupling*: where a node in one layer is connected to itself in adjacent layers. In other words $(v, v, l, k) \in C$ if $|l - k| == 1$. This is the default coupling when using multiplex networks to model dynamic networks.
- *Categorical coupling*: where a node in one layer in connected to itself in each other layer. This is the default coupling when representing multi-relationnal networks.

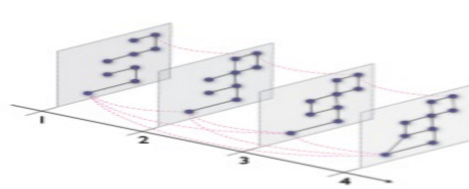

Fig. 2. Ordinal and categorical couplings

Other coupling schemes can also be considered as discussed in [10].

Typically, one first operation to apply to multiplex networks is *network flatten*. This consists on transforming the multiplex into a monoplex network. The goal is to have a baseline for comparing different multiplex networks. This can allow applying classical network analysis tools to multiplex ones. Flattened network is obtained by applying a *layer aggregation* function. In general, a layer aggregation approach transforms a multiplex network into a weighted monoplex graph : $G =< V, E, W >$ where W is a weight matrix. Different weights computations approaches can be applied. One simple aggregation function is the binary weighting: two nodes u, v are linked in the aggregated simple graph if there is at least one layer in the multiplex where these nodes are linked. Formally we have:

$$w_{ij} = \begin{cases} 1 & \text{if } \exists 1 \leq i \leq \alpha : (i,j) \in E_i \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Other layer aggregation functions have been also proposed [11], [12], [13]. Whatever is the applied aggregation function, the aggregation process will lead to information loss since different types of links will treated indifferently. In [14] authors explore to which extent some layers of a multiplex network can be merged without information loss. Table I mains notations used later in this paper.

TABLE I
MULTIPLEX NETWORKS: NOTATIONS

Notation	Description
$A^{[k]}$	Slice k Adjacency matrix
$d_i^{[k]}$	Degree of node i in slice k
$d_i^{tot} = \sum_{s=1}^{\alpha} d_i^{[s]}$	Total degree of node i
$m^{[k]}$	edge number in slice k
$\Gamma(v)^{[k]} = \{u \in V : (u,v) \in E_k\}$	Neighbor's of v in slice k
$\Gamma(v)^{tot} = \cup_{s \in \{1,\ldots,\alpha\}} \Gamma(v)^{[s]}$	Neighbors of v in all α slices
$SPath^{[k]}(u,v)$	Shortest path length between nodes u and v in slice k

III. CENTRALITIES & DYADIC METRICS

Computing basic centralities (degree, proximity, betweenness, etc.) requires first defining basic concepts such as node's degree, node's neighborhood and shortest paths in multiplex networks [3]. We discuss these basic issues in next paragraphs.

Neighborhood: Different options can be considered to define the neighborhood of a node in a multiplex. One simple approach is to make the union of all neighbors across all layers. Another more restrictive definition is to compute the intersection of node's neighbors sets across all layers. In [3], [15], authors define a multiplex neighborhood of a node by introducing a threshold on the number of layers in which two nodes are linked. Formally we have:

$$\Gamma_m(v) = \{u \in V \text{ such that } count(i) > m : A_{vy}^{[i]} > 0\}$$

We extend further this definition by proposing a similarity-guided neighborhood: Neighbors of a node v are computed as a subset of $\Gamma(v)^{tot}$ composed of nodes having a similarity with v exceeding a given threshold δ. using the classical Jaccard similarity function this can be formally written as follows:

$$\Gamma^{mux}(v) = \{x \in \Gamma(v)^{tot} : \frac{\Gamma(v)^{tot} \cap \Gamma(x)^{tot}}{\Gamma(v)^{tot} \cup \Gamma(x)^{tot}} \geq \delta\} \quad (2)$$

$\delta \in [0, 1]$ is the applied threshold.

The threshold δ allow to fine-tune the neighborhood size ranging from the most restrictive definition (interaction of neighborhood sets across all layers) to the most loose definition (the union of all neighbors across all layers).

Node degree: The degree of a node is defined as the cardinality of the set of direct neighbors. By defining the multiplex neighborhood function we can define directly a multiplex node degree function. Another interesting multiplex degree function has been proposed in [6]. It defines the multiplex degree of a node as the entropy of node's degrees in each layer. In a formal way we can write:

$$d_i^{multiplex} = -\sum_{k=1}^{\alpha} \frac{d_i^{[k]}}{d_i^{[tot]}} log\left(\frac{d_i^{[k]}}{d_i^{[tot]}}\right) \quad (3)$$

The basic idea underlying this proposition, is that a node should be involved in more than one layer in order to qualify; otherwise its value is zero. The degree of a node i is null if all its neighbors are concentrated in a single layer. However, it reaches its maximum value if the number of neighbors is the same in all layers. This can be useful if we have no prior information about the importance of each layer in the studied multiplex but we want to stress that all layers are important to the target analysis task.

Shortest path: Two approaches can be applied to compute the length of the shortest-path between two nodes in a multiplex network: The first approach consist in computing the shortest-path in an aggregated network. The second approach consists in computing an aggregation of the shortest path lengths across all layers.

IV. COMMUNITY DETECTION

In real-world complex networks nodes are generally arranged in tightly knit groups that are loosely connected one to each other. Such groups are called communities. Community members are generally admitted to share common proprieties. Hence, unfolding the community structure of a network could give us many insights about the overall structure of the network. This problem has attracted much of attention in past years. Most of existing approaches are designed for simple networks, where all edges are of the same type [16]. Different approaches have been recently proposed to cope with this problem in the context of multiplex networks [8]. We can classify existing approaches into two broad classes:

1) *Applying monoplex approaches*: the basic idea is to transform the problem into a problem of community detection in simple networks [17], [18].
2) Extending existing algorithms to deal directly with multiplex networks [19], [20].

Next we detail both approaches.

A. Applying monoplex approaches

One first approach consists appling layer aggregation approaches and then apply classical communjity detection on the flatten network [12], [17]. In [21] an original multiplex transformation approach is proposed. It consists of mapping a multiplex to a 3-uniform hyper-graph $H = (V^*, E^*)$ such that the node set in the hyper-graph is $V^* = V \cup 1, ..., \alpha$ and $(u, v, i) \in E^* if \exists l : A^l_{uv} \neq 0, u, v \in V, i \in 1, ..., \alpha$. Community detection algorithms in hyper-graphs can then be applied on the obtained graph. In [20] a multi-objective approach is applied. The idea is to apply a classical community detection algorithm to a first layer. For all consecutive layers a bi-objective optimization approach is applied in order to detect communities that maximize both the modularity in the current layer and the similarity to the community structure detected on the previous layer. This approach can be applied to multiplex networks where ordinal coupling is applied.

Another way is to to apply a classical community detection to each layer of a multiplex then merging obtained partitions using ensemble clustering approaches [22].

B. Extending monoplex approaches to the multiplex case

Few studies have addressed the problem of simultaneous exploration of all layers of a multiplex network for the detection of communities. [11] is among the first studies that have tried to extend existing approaches to multiplex setting. The leading role that modularity and its optimization have played in the context of community detection in simple graphs has naturally motivated works to generalize the modularity to the case of multiplex networks. A generalized modularity function is proposed in [23]. This is given as:

$$Q_{multiplex}(P) = \frac{1}{2\mu} \sum_{c \in P} \sum_{\substack{i,j \in c \\ k,l:1 \to \alpha}} \left(\left(A^{[k]}_{ij} - \lambda_k \frac{d^{[k]}_i d^{[k]}_j}{2m^{[k]}} \right) \right) \tag{4}$$

Where $\mu = \sum_{k:1 \to \alpha} m^{[k]}$ is a normalization factor, and λ_k is a resolution factor as introduced [24] in order to cope with the modularity resolution problem. Approaches based on optimizing the multiplex modularity are likely to have the same drawbacks of those optimizing the original modularity function for monoplex approaches [25]. This motivates exploring other approaches for community detection. The Infomap algorithm [26] has also been extended to the multiplex case [27]. In [28] an adaptation of the *Walktrap* community detection algorithm [29] is proposed. A seed-centric approach is also proposed in [30].

C. Evaluation criteria

The problem of evaluating community detection algorithm still to be an open problem despite the great amount of work conducted in this field [31]. Since few multiplex networks with ground truth partitions are available, *unsupervised* evaluation metrics are generally used. These include the multiplex modularity ($\mathcal{Q}$) (see 4), the redundancy (ρ) criteria and the complementarity (γ) criteria introduced in [12].

Redundancy criteria (ρ) **[12] :** The redundancy ρ computes the average of the redundant link of each intra-community in all multiplex layers. The intuition is that the link intra-community should be recurring in different layers. The computing of this indicator is as follows: We denote by:

- P the set of couple (u, v) which are directly connected to at least one layer.
- $\bar{\bar{P}}$ the set of couple (u, v) which are directly connected in at least two layers.
- $P_c \subset P$ represents all links in the community c
- $\bar{\bar{P}}_c \subset \bar{\bar{P}}$ the subset of $\bar{\bar{P}}$ and which are also in c.

The redundancy of the community c is given by:

$$\rho(c) = \sum_{(u,v) \in \bar{\bar{P}}_c} \frac{\| \{k : \exists A^{[k]}_{uv} \neq 0\} \|}{\alpha \times \| P_c \|} \tag{5}$$

The quality of a given multiplex partition is defined as follow:

$$\rho(\mathcal{P}) = \frac{1}{\| \mathcal{P} \|} \sum_{c \in \mathcal{P}} \rho(c) \tag{6}$$

$$\gamma(P) = \frac{1}{\|P\|} \sum_{c \in P} \gamma(c) \tag{7}$$

Complementarity criteria (γ) **[12] :** The complementarity γ is the conjunction of three measures :

- Variety $\mathcal{V}_c$: this is the proportion of occurrence of the community c across layers of the multiplex.

$$\mathcal{V}_c = \sum_{s=1}^{\alpha} \frac{\|\exists (i,j) \in c / A^{[s]}_{ij} \neq 0\|}{\alpha - 1} \tag{8}$$

- Exclusivity ε_c : this is the number of pairs of nodes, in community c, that are connected exclusively in one layer.

$$\varepsilon_c = \sum_{s=1}^{\alpha} \frac{\|\overline{P_{c,s}}\|}{\|P_c\|} \tag{9}$$

  with P_c : is the set of pairs (i, j) in community c that are connected at least in one layer. $\overline{P_{c,s}}$: is the set of pairs (i, j) in community c that are connected exclusively in layer s.

- Homogeneity $\mathcal{H}_c$: this captures how uniform is the distribution of the number of edges, in the community c, per layer. The idea is that intra-community links must have a uniform distribution among all layers.

$$\mathcal{H}_c = \begin{cases} 1 & if \quad \sigma_c = 0 \\ 1 - \frac{\sigma_c}{\sigma^{max}_c} & otherwise \end{cases} \tag{10}$$

with

$$avg_c = \sum_{s=1}^{\alpha} \frac{\|P_{c,s}\|}{\alpha}$$

$$\sigma_c = \sqrt{\sum_{s=1}^{\alpha} \frac{(\|P_{c,s}\| - avg_c)^2}{\alpha}}$$

$$\sigma^{max}_c = \sqrt{\frac{(max(\| P_{c,d} \|) - min(\| P_{c,d} \|))^2}{2}}$$

The higher the complementarity the better is the partition. The complementarity is then given by the following formula:

$$\gamma(c) = \mathcal{V}_c \times \varepsilon_c \times \mathcal{H}_c$$

V. Multiplex analysis tools

Recently, two different packages for multiplex network analysis have been released: *muxviz* [5] and *muna* [32]. The first is an R package that focuses mainly on multiplex network visualisation (see figure 1). It provides also a support for implementing some layer-aggregation approaches and implements generalized modularity–based community detection algorithm. The second package *Muna*, is provided as an extension of the *igraph* graph analysis API [33]. It is provided in two versions *R* and *Python* as is provided under GPL licence and can be downloaded from http://lipn.fr/~kanawati/software. A special attention in *Muna* is made to the problem of community detection and evaluation in multiplex networks. It actually provides an extensive set of different community detection and evaluation approaches.

VI. Conclusion

The curent maturity of *network science* coupled with the availability of huge amount of heterogeneous data in different fields allow today a move to a more complex representations of real-world interactions. The multiplex network model is one promising option. It is powerful enough to model multi relational, dynamic and attributed networks. Therefore this model is attracting an increasing attention form different researchers from different communities. In this paper we have provided a quick survey of recent advances in the field of multiplex network analysis and mining.

References

[1] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of complex networks*, vol. 2, pp. 203–271, 2014. [Online]. Available: http://arxiv.org/abs/1309.7233

[2] M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, and D. Pedreschi, "Multidimensional networks: foundations of structural analysis," *World Wide Web*, vol. 16, no. 5-6, pp. 567–593, 2013.

[3] P. Brodka and P. Kazienko, *Encyclopedia of Social Network Analysis and Mining*. Springer, 2014, ch. Multi-layered social networks.

[4] E. Lazega, *The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership*. Oxford university Press, 2001.

[5] M. D. Domenico, M. A. Porter, and A. Arenas, "Multilayer analysis and visualization of networks," *J. Complex Netw. (2014)*, vol. 10, 2014. [Online]. Available: http://arxiv.org/abs/1405.0843

[6] F. Battiston, V. Nicosia, and V. Latora, "Metrics for the analysis of multiplex networks," *CoRR*, vol. abs/1308.3182, 2013.

[7] M. Magnani and M. Marzolla, "Path-based and whole-network measures," in *Encyclopedia of Social Network Analysis and Mining*, 2014, pp. 1256–1269. [Online]. Available: http://dx.doi.org/10.1007/978-1-4614-6170-8_241

[8] C. W. Loe and H. Jeldtoft Jensen, "Comparison of communities detection algorithms for multiplex," *ArXiv e-prints*, Jun. 2014.

[9] M. Pujari and R. Kanawati, "Link prediction in multiplex networks," *Networks and Heterogeneous Media*, vol. 10, pp. 17–35, March 2015, special Issue on New trends, models and applications in Complex and Multiplex Networks.

[10] T. Murata, "Comparison of inter-layer couplings of multilayer networks," in *The 4th International Workshop on Complex Networks and their Applications*, Bangkok, Thailand, November 2015.

[11] L. Tang and H. Liu, *Community Detection and Mining in Social Media*, ser. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers, 2010.

[12] M. Berlingerio, M. Coscia, and F. Giannotti, "Finding and characterizing communities in multidimensional networks," in *ASONAM*. IEEE Computer Society, 2011, pp. 490–494.

[13] D. Cai, Z. Shao, X. He, X. Yan, and J. Han, "Mining hidden community in heterogeneous social networks," in *ACM-SIGKDD Workshop on Link Discovery: Issues, Approaches and Applications (LinkKDD'05)*, Chicago, IL, Aug 2005.

[14] M. D. Domenico, V. Nicosia, A. Arenas, and V. Lator, "Structural reducibility of multilayer networks," *Nature communications*, vol. 6, p. 6864, 2015.

[15] P. Kazienko, P. Brodkadka, and K. Musial, "Individual neighbourhood exploration in complex multi-layered social network," in *Web Intelligence/IAT Workshops*, 2010, pp. 5–8.

[16] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[17] D. D. Suthers, J. Fusco, P. K. Schank, K.-H. Chu, and M. S. Schlager, "Discovery of community structures in a heterogeneous professional online network," in *HICSS*. IEEE, 2013, pp. 3262–3271.

[18] M. Berlingerio, F. Pinelli, and F. Calabrese, "Abacus: frequent pattern mining-based community discovery in multidimensional networks," *Data Min. Knowl. Discov.*, vol. 27, no. 3, pp. 294–320, 2013.

[19] R. Lambiotte, "Multi-scale modularity in complex networks," in *WiOpt*. IEEE, 2010, pp. 546–553.

[20] A. Amelio and C. Pizzuti, "Community detection in multidimensional networks," in *IEEE 26th International Conference on Tools with Artificial Intelligence*, 2014, pp. 352–359.

[21] P. M. Comar, P.-N. Tan, and A. K. Jain, "Simultaneous classification and community detection on heterogeneous network data," *Data Min. Knowl. Discov.*, vol. 25, no. 3, pp. 420–449, 2012.

[22] R. Kanawati, "Community detection in social networks: the power of ensemble methods." in *ACM/IEEE International conference on data science & advanced analytics*. Shanghai: IEEE, November 2014.

[23] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *Science*, vol. 328, no. 5980, pp. 876–878, 2010.

[24] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection." *Physical Review E*, vol. 74, no. 1, 2006.

[25] B. H. Good, Y.-A. de Montjoye, and A. Clauset., "The performance of modularity maximization in practical contexts." *Physical Review*, vol. E, no. 81, p. 046106, 2010.

[26] M. Rosvall, D. Axelsson, and C. T. Bergstrom, "The map equation," *Eur. Phys. J. Special Topics*, vol. 13, p. 178, 2009.

[27] M. D. Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, "Identifying modular flows on multilayer networks reveals highly overlapping organization in social systems," *Phys. Rev*, vol. 5, p. 011027, 2015.

[28] Z. Kuncheva and G. Montana, "Community detection in multiplex networks using locally adaptive random walks," in *MANEM 2workshop - Proceedings of ASONAM 2015*, Paris, August 2015.

[29] P. Pons and M. Latapy, "Computing communities in large networks using random walks," *J. Graph Algorithms Appl.*, vol. 10, no. 2, pp. 191–218, 2006.

[30] M. Hmimida and R. Kanawati, "Community detection in multiplex networks: A seed-centric approach," *Networks and Heterogeneous Media*, vol. 10, no. 1, pp. 71–85, March 2015, special Issue on New trends, models and applications in Complex and Multiplex Networks.

[31] Z. Yakoubi and R. Kanawati, "Licod: Leader-driven approaches for community detection," *Vietnam Journal of Computer Science*, vol. 1, no. 4, pp. 241–256, 2014.

[32] I. Falih and R. Kanawati, "Muna: A multiplex network analysis library," in *The 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Paris, August 2015, pp. 757–760.

[33] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, p. 1695, 2006. [Online]. Available: http://igraph.sf.net