NodeTrix-Multiplex: Visual Analytics of Multiplex Small World Networks

Shivam Agarwal, Amit Tomar and Jaya Sreevalsan-Nair

Abstract

Analyzing multiplex small world networks (SWNs) using community detection (CD) is a challenging task. We propose the use of visual analytics to probe and extract communities in such networks, where one of the layers defines the network topology and exhibits small-world property. Our novel visual analytics framework, NodeTrix-Multiplex (NTM), for visual exploration of multiplex SWNs, integrates focus+context network visualization, and analysis of community detection results, within the focus. We propose a heterogeneous data model, which composites multiple layers for the focus and context and thus, enables finding communities across layers. We perform a case-study on a co-authorship (collaboration) network, with a functional layer obtained from the author-topic similarity graph. We also perform an expert user evaluation of the tool, developed using NTM.

I. Introduction

Complex networks are real-world, ubiquitous and important, as networks can simultaneously encode objects in a specific context and the pairwise relationships between those objects. Small world networks (SWNs) are a class of complex networks [1], [31], which shows small-world property. Social networks, such as collaboration networks, are SWNs. Owing to the advances in technological capability of gathering, storing, and analysis of these data sets, such networks are increasingly encoding more information. Thus, the rich data is stored as multiplex complex networks, where different relationships, between the same set of nodes, are stored as separate *layers*. The layers of the multiplex network have unique adjacency matrices [3], [15]. Since our focus is on multiplex SWNs, we assume one of the layers in the network gives the network topology of a SWN, which in turn determines an initial community formation. We call such a layer "*structural*" layer, and the other layers, such as similarity graphs, "*functional*" layers, borrowing terminology from brain networks [19]. Another way to look at it is that, we use the *existential* layer (i.e. the layer that has caused the very existence of the complex network) as the structural network, e.g. collaboration network. Thus, the other layers are "functional," which depend on the existenial layer. In the case of multiplex SWNs, we consider the existential layer, that exhibits the small-world property, to be the structural one.

Community detection (CD) can reveal several patterns in a complex network. However, CD across multiple layers is challenging owing to the differences in *"percolation"* of communities in the layers [8]. Here, we focus in selectively exploring the dynamics of communities within a small subnetwork in the complex network, which is a community in itself. Thus, for community exploration and detection in multiplex SWNs, we propose a **focus+context paradigm**, and a visual analytic framework, **NodeTrix-Multiplex** (NTM), that enables the user to see clustering tendencies in the focus. Visual analytics is an active area of research where visualization plays a larger role in data analytics, in an interweaved manner, than just summarizing information or exploring data. Figure 1 summarizes our proposed work, which shows **our proposed heterogeneous data model** (HDM), on which visual analytics is used for *drilling down* across layers in a subnetwork of interest. Our proposed visual analytic framework is designed with the visual information seeking mantra: *overview first, zoom and filter, then details on demand* [26]. NTM uses the hybrid visual representation of SWNs, as proposed in NodeTrix [12], which exploits the "locally dense, globally sparse" structure of a SWN. A preliminary version of our tool[1] is available at http://nmultiplex.au-syd.mybluemix.net/

Notations:: We denote a multiplex network with N layers (each defined by a unique adjacency matrix), as $\mathscr{M} = \{\mathscr{V}(\mathscr{M}), \mathscr{E}^0, \ldots, \mathscr{E}^{N-1}\}$, where $\mathscr{V}(\mathscr{M})$ is the vertex set of the network, and $\mathscr{E}^i$ is the set of edges belonging to the i^{th} layer, and it is represented by the weighted adjacency matrix of the i^{th} layer. $e(u,v)$ implies an edge exists between vertices $u,v \in \mathscr{V}$ and it encodes the edge weight, a real value.

The i^{th} layer of $\mathscr{M}$ is defined as $\mathscr{L}^i = \{\mathscr{V}(\mathscr{M}), \mathscr{E}^i\}$. Non-overlapping (or crisp) communities in any layer $\mathscr{L}^i$, are denoted as $\{\mathscr{C}^i_0, \ldots, \mathscr{C}^i_{M_i-1}\}$ for M_i communities, where $\mathscr{C}^i_j$ is the vertex set of the j^{th} community in the i^{th} layer. Thus, $0 \leq i < N$ and $0 \leq j,k < M_i$ where $j \neq k$, we get $\mathscr{V}(\mathscr{C}^i_j) \subset \mathscr{V}(\mathscr{M})$ and $\mathscr{V}(\mathscr{C}^i_j) \cap \mathscr{V}(\mathscr{C}^i_k) = \emptyset$.

Any subnetwork in $\mathscr{L}^k$ is given as $\mathscr{N}(k)$, where its vertex set is $\mathscr{V}(\mathscr{N}(k)) \subset \mathscr{V}(\mathscr{M})$, and its edge set is $E(\mathscr{N}(k)) = \{e(u,v)\}|u,v \in \mathscr{V}(\mathscr{N}(k)) \wedge e(u,v) \in \mathscr{E}^k\}$. However, a subnetwork in $\mathscr{L}^k$ can be constructed using the vertex set of community $\mathscr{C}^i_j$, where $i \gtreqless k$; in which case, the subnetwork is given as: $\mathscr{N}(k, \mathscr{C}^i_j)$, whose vertex set is $\mathscr{V}(\mathscr{C}^i_j)$ and edge set is $E(\mathscr{N}(k, \mathscr{C}^i_j)) = \{e(u,v)|u,v \in \mathscr{V}(\mathscr{C}^i_j) \wedge (e(u,v) \in \mathscr{E}^k)\}$.

This document is preprint, as on October 17, 2016. The authors are with Graphics-Visualization-Computing Lab, IIIT Bangalore. email: {shivam.agarwal,amit.tomar}@iiitb.org , jnair@iiitb.ac.in

[1]The tool is best readable on the Chromium browser.

Our proposed focus and context exist in $\mathscr{L}^k$ and pertain to a subnetwork $\mathscr{N}(k)$, and hence, are denoted as $\mathscr{F}(\mathscr{N}(k))$ and $\mathscr{U}(\mathscr{N}(k))$. The shorthand notations for vertex sets of focus and context are V_F and V_U, respectively; and the edge sets are E_F and E_U, respectively. Even though interchangeably used as synonyms, here, we use "network", "multiplex network","nodes" and "links" in the context of dataset, and "graph", "multigraph", "vertices", and "edges" as data structures, respectively.

II. Related Work

In our work, visualizing communities within a SWN and exploring them are key ideas. Prior to visualizing, we detect communities using state-of-the-art algorithms; and for exploring the communities, we use matrix seriation. While there is not much material on visualization of multiplex networks, CD in multiplex networks has been an active area of research. Notwithstanding, as SWNs is a class of complex networks, here, we discuss relevant literature in complex networks as well.

Visualization of Communities in Complex Networks:: NodeTrix [12], is a visualization of social networks, where the small-world property of "globally sparse but locally dense" has been exploited to provide the visual representation, which integrates better readability of node-link and matrix representations of the network in respective scenarios (i.e. sparse and dense nature of the network which in the global and local spatial context, respectively) [11]. The locally dense subgraphs are represented as *"aggregated nodes"* (ANs), and rendered as matrices. We direct the readers to the state of the art article on visualizations of groups in graphs [29]. Node-link diagrams and integrated (linked) views have been widely used for visualizing hierarchical structures in networks [24], [25], [30], and for multivariate networks [28], [14], [17]. Bastian et al. [2] have proposed Gephi, a popular network visualization tool, which shows connected components and communities using node-link diagram.

Community Detection in Complex Networks:: Modularity-based Louvain CD [7] and graph-theoretic based Tarjan's algorithms [27] are popularly used for extracting communities and strongly connected components in networks, respectively. Algorithms for hierarchical CD in multiplex networks, for finding crisp communities, use modularity across layers/slices as a guiding principle [5], [18], to determine the best community formation. While these algorithms have composited layers in the multiplex network at the node-level, we propose to perform the same at a coarser level of granularity, i.e. we composite communities, or subnetworks; to make it more scalable for interactive visualizations. de Domenico et al. [10] have proposed the use of modular flows between nodes across layers to identify overlapping communities in multilayer networks. We use a similar concept, except that de Domenico et al. have proposed modular flows across several layers in communities, whereas ours pertain to "modular flows" in aggregated nodes (as used in NodeTrix) across layers in multiplex networks. There have been several studies on visual analytics of multiplex networks such as, Renoust et al. [21] and Rossi and Magnani [23], that have discussed the limitations of extending simplex network visualizations to multiplex ones. They have worked with each network "slice" or layer having its own independent graph layout. As opposed to their work which focuses on visual analytics of dynamics across layers using node-link diagrams predominantly, our work is on CD across layers using a hybrid visualization. Our visualization is however biased towards the SWN layer, owing to which we do not compute layouts for other layers.

Matrix Seriation:: Seriation is a process of reordering rows or columns in a matrix to identify pertinent patterns of clustering. Visual assessment of clustering tendency (VAT) algorithm [6] computes the minimum spanning tree of the dissimilarity graph to give ordering of nodes, and upon reordering, the clusters show the pattern of square blocks along the diagonal of the matrix. Parveen et al. [20] have demonstrated that similarity matrices, after automatic seriation using VAT algorithm, can provide effective matrix visualization of SWNs. We direct the readers to surveys of matrix reordering methods for different domains [16] and for network visualization [4].

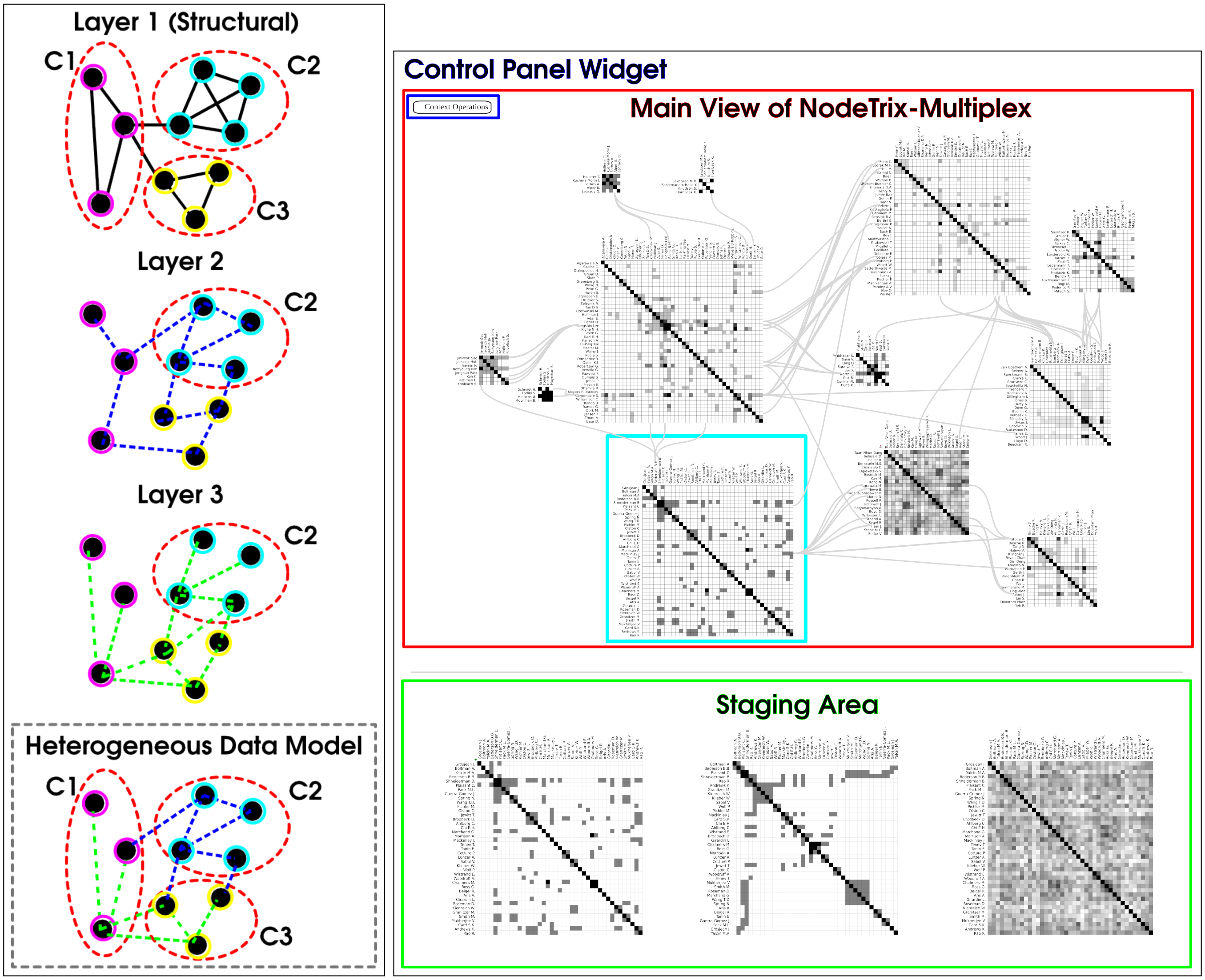
III. Focus+Context Approach and Data Model

We propose a focus+context paradigm to probe communities in a subnetwork of interest within the multiplex network. Since we are interested in studying multiple layers of the complex network, our paradigm must be integrated with a HDM. Our rationale is that the focus, which is a subnetwork, will allow us to study localized trends of the network. At the same time, the focus has to be studied in the presence of context, for which we use the rest of the network. In our work, we propose to use a subnetwork ($\mathscr{N}(k)$) in a specific layer ($\mathscr{L}^k$) as the focus ($\mathscr{F}(\mathscr{N}(k))$); thus, the remaining network becomes the context ($\mathscr{U}(\mathscr{N}(k))$). The vertex and edge sets for the focus (V_F and E_F) and context (V_U and E_U) are:

$$\begin{aligned} V_F &= \mathscr{V}(\mathscr{F}(\mathscr{N}(k))) = \mathscr{V}(\mathscr{N}(k)); \\ E_F &= E(\mathscr{F}(\mathscr{N}(k))) = E(\mathscr{N}(k)) \cup \{e(u,v) | (u \in V_F \wedge v \in V_U \wedge e(u,v) \in \mathscr{E}^k) \vee \\ &\qquad (u \in V_U \wedge v \in V_F \wedge e(u,v) \in \mathscr{E}^k)\}; \end{aligned}$$

$$V_U = \mathscr{V}(\mathscr{U}(\mathscr{N}(k))) = \mathscr{V}(\mathscr{M}) \setminus V_F; \; E_U = E(\mathscr{U}(\mathscr{N}(k))) = \mathscr{E}^k \setminus E_F. \tag{1}$$

In order to find a *subnetwork of interest*, we propose to perform CD in the concerned layer $\mathscr{L}^k$, thus getting M_k non-overlapping communities $\mathscr{C}_0^k, \dots, \mathscr{C}_{M_k-1}^k$; and then, use one of the communities as a *subnetwork of interest*. Thus, one such community is treated as the focus, and the remaining network becomes the context. Thus, V_F, E_F, V_U, E_U in Equation 1 can now be written as: $\mathscr{V}(\mathscr{F}(\mathscr{N}(k,\mathscr{C}_j^k)))$, $E(\mathscr{F}(\mathscr{N}(k,\mathscr{C}_j^k)))$, $\mathscr{V}(\mathscr{U}(\mathscr{N}(k,\mathscr{C}_j^k)))$ and $E(\mathscr{U}(\mathscr{N}(k,\mathscr{C}_j^k)))$, respectively.



(a) (b)

Fig. 1: (a) Schematic of our HDM for multiplex network with three layers, the structural layer (Layer 1) and two functional layers (Layers 2 and 3). Of the communities C1, C2, C3 in Layer 1, the intra- and inter-community edges of the focus (i.e. C2) can be taken from Layer 2 [blue dashed lines]; and those of the context from Layer 3 [green dashed lines]. (b) GUI layout of NTM shows the *main view* [red], widget for expanding the *control panel* [blue] and the *staging area* [green]. A subnetwork of IV dataset (233 nodes, 569 edges, 12 different communities/ANs), with the co-authorship layer in both ANs and links is displayed in the main view. Images of the focus/AN [cyan] from the main view are saved in its staging area; showing (left-to-right) unseriated co-authorship layer, VAT-seriated co-authorship layer, and VAT-seriated author-topic similarity layer.

Using the aforementioned construction of focus, the communities and the focus+context paradigm lie in the same layer, and hence, this pertains to analysis of a single-layer network. What if we use the community in one layer to define the focus, which is further studied across multiple layers in a multiplex network ?

> There is a subtle difference between our usage of terms, *"community"* and *"focus"*. The edge set of the former consists of the intra-community edges exclusively; whereas that of the latter (E_F, as used in Equation 1) is the set of all edges (both intra-community edges and inter-community), for which at least one of the vertices belong to the community.

Heterogeneous Data Model:: For a multiplex network, we propose the construction of a *composited single-layer network* $\mathscr{M}_{mod}$, which is an aggregate of multiple network layers. Our proposed algorithm, of $\mathscr{O}(|\mathscr{V}(\mathscr{M})|)$ complexity, aggregates a maximum of three layers of $\mathscr{M}$, taken at a time, in a three-step process (Figure 1(a)). Firstly, we perform CD in layer $\mathscr{L}^i$ to find *subnetwork of interest* $\mathscr{C}^i_j$. Secondly, using the vertex set $V_F = \mathscr{V}(\mathscr{C}^i_j)$ in layer $\mathscr{L}^k$ we construct *focus*, $\mathscr{F}(\mathscr{N}(k, \mathscr{C}^i_j))$.

Thirdly, we define *context*, $\mathscr{U}(\mathscr{N}(u,\mathscr{C}_j^i))$, using vertex set, $V_U = \mathscr{V}(\mathscr{M}) \setminus V_F$, but edge set from a third layer $\mathscr{L}^u$. Since, we are able to reconstruct a single *"composite"* layer using multiple layers, we call this construction a *heterogeneous data model*. Thus, rewriting Equation 1 for multiple layers:

$$\begin{aligned} E_F &= E(\mathscr{F}(\mathscr{N}(k,\mathscr{C}_j^i))) = E(\mathscr{N}(k)) \cup \{e(u,v)|(u \in V_F \wedge v \in V_U \wedge e(u,v) \in \mathscr{E}^k) \vee \\ &\qquad\qquad (u \in V_U \wedge v \in V_F \wedge e(u,v) \in \mathscr{E}^k)\}; \\ E_U &= E(\mathscr{U}(\mathscr{N}(u,\mathscr{C}_j^i))) = \mathscr{E}^u \setminus \{e(u,v)|(u \in V_F) \vee (v \in V_F)\} \end{aligned} \tag{2}$$

Our rationale is that we can *switch between different layers* in the focus and context and study *localized patterns*, such as in CD, persistent across the layers.

Since, in our case, the structural layer exhibits the small-world property and contains *"locally dense"* subnetworks, we perform CD in $\mathscr{L}^0$. The sparse links between these communities in $\mathscr{L}^0$ also indicate that the communities internally are well-connected, which implies analysis of each of these communities can be performed mostly independently. Hence, owing to the better defined community formation in $\mathscr{L}^0$, our analysis and graph layout are more biased to it than to the other layers. We use one such community in $\mathscr{L}^0$ as the focus. We find: $V_F = \mathscr{V}(\mathscr{C}_j^0)$; $V_U = \mathscr{V}(\mathscr{M}) \setminus V_F$; $E_F = E(\mathscr{F}(\mathscr{N}(k,\mathscr{C}_j^0)))$; and $E_U = E(\mathscr{U}(\mathscr{N}(u,\mathscr{C}_j^0)))$. This model can be generically used for two-layer multiplex network, where one of the two layers can be treated as $\mathscr{L}^u$, as done in our case-study.

IV. NodeTrix-Multiplex: A Visual Analytic Framework

We propose NodeTrix-Multiplex (NTM), which is a visual analytic framework built on the concepts and visualization layout used in NodeTrix [12]. NTM is a human-in-the-loop framework, which enables users to visually explore and find strong communities which *percolates* across layers of a multiplex SWN. It is integrated with our HDM, which uses focus+context paradigm and a seriation algorithm. It enables the user to understand the dynamics of community formation in different layers by drilling down a subnetwork of interest. The choice of using NodeTrix over node-link diagrams, e.g. in Gephi [2], is due to clear separability of the matrix visualization of focus from the context, in the former (Figure 2). This separability helps in visualization of composited network layer, using different layers for CD, the focus, and the context (Figure 1(a)).

GUI Layout and User Interactions:: The proposed layout of GUI for NTM (Figure 1(b)) consists of three components: main view, staging area, and control panel. The hybrid visualization of the focus+context is shown in the **main view**, where the user can choose a focus. The user can interact with the focus and context simultaneously or exclusively with either. In the **staging area** the user can save images of the focus and view them in different zoom levels. In **the control panel**, the user has the controls to choose the layer for focus/ context visualization, threshold for ε-neighborhood for similarity graph (i.e., if a similarity layer is present in the network), color scheme for colormapping of matrices, and seriation. These operations are for the focus and its context, which can be applied simultaneously or exclusively to either, using locking of focus. Separate choices of layer for the focus and the context support the HDM (Section III) and VAT seriation for the focus (Section II).

Key Differences between NodeTrix and NTM::

1) NodeTrix is exclusively for studying all ANs in a single-layer SWN homogeneously; whereas our goal is to study local trends in the the multiplex SWN heterogeneously. Our heterogeneous study implies studying an AN in settings different from those of other nodes/ ANs in the network.
2) Owing to the difference in the motivation, NodeTrix uses user-guided agglomeration to create ANs, whereas we use Louvain CD algorithm [7] to automatically extract strong communities in the structural layer. The communities are represented as ANs in NTM.
3) NodeTrix uses user-guided seriation for finding patterns in matrices, whereas we use automatic seriation algorithm, such as VAT algorithm [6].
4) NodeTrix visualizes unweighted adjacency matrix, whereas NTM uses weighted adjacency matrices, for CD, and their complements, i.e. distance matrices, for visualization. The latter is done to comply with the visualization used in VAT algorithm. The difference is that the diagonal cells of AN have value one in NodeTrix (colored white) and value zero in NTM (colored black).
5) The visualization tasks are different – the tasks in both NodeTrix and NTM are to identify communities (T1), central actors (T2), and roles and positions (T3); and NTM additionally has to analyze CD across layers. NTM accomplishes T1 without visual interaction. For T2 and T3, VAT seriation of ANs in NTM highlights the cross, block, and intermediate pattern, as in [12]. The additional unique tasks for NTM are: (T4) find a set of nodes in a community which show clustering tendency across different layers, using the focus, and (T5) find inter-community relationships which could be strong in layers different from the one used for CD, using focus+context.

Figure[2] 1 shows the layout of the GUI. In the main view, the user can move matrices of the aggregated nodes, which updates the links between the ANs. The operations, which are facilitated through the control panel of NTM, are implemented on both the focus as well as the context. Additionally, depending on the user's needs, these operations can be implemented separately, for which we introduce the notion of "locking" the focus, to preserve it from the modifications made to the context. Thus,

[2]All images in this paper look best when zoomed in.

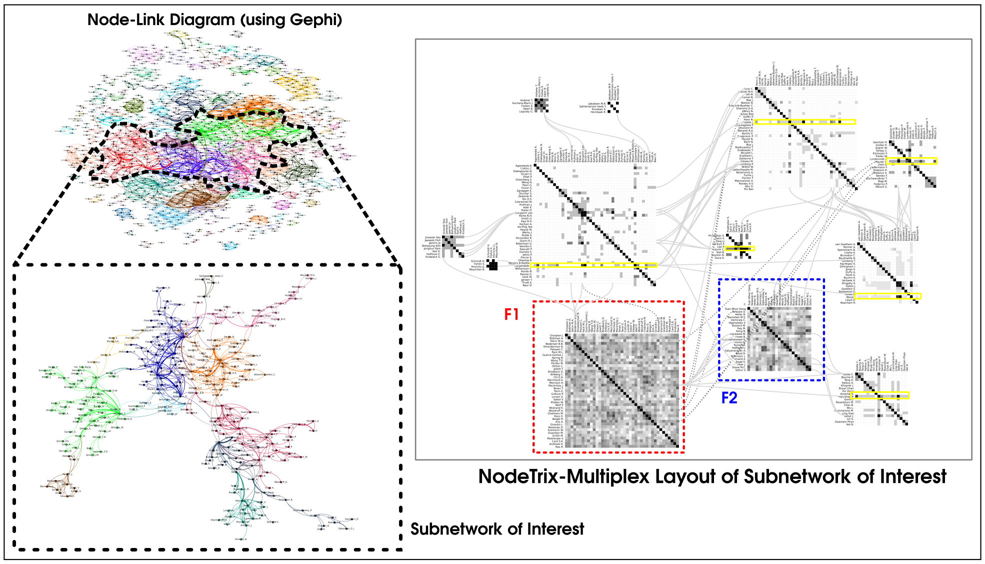


Fig. 2: A subnetwork (233 nodes, 569 edges, and 12 ANs/communities) in the IV co-authorship network dataset shows the foci, F1 and F2, in the author-topic similarity graph (functional layer) and context in the co-authorship layer (structural layer). Yellow highlights show central actors in the community/AN. The inter-community edges are shown in both functional [dotted lines, showing 22 edges with similarity score > 0.7] and structural [solid lines] layers.

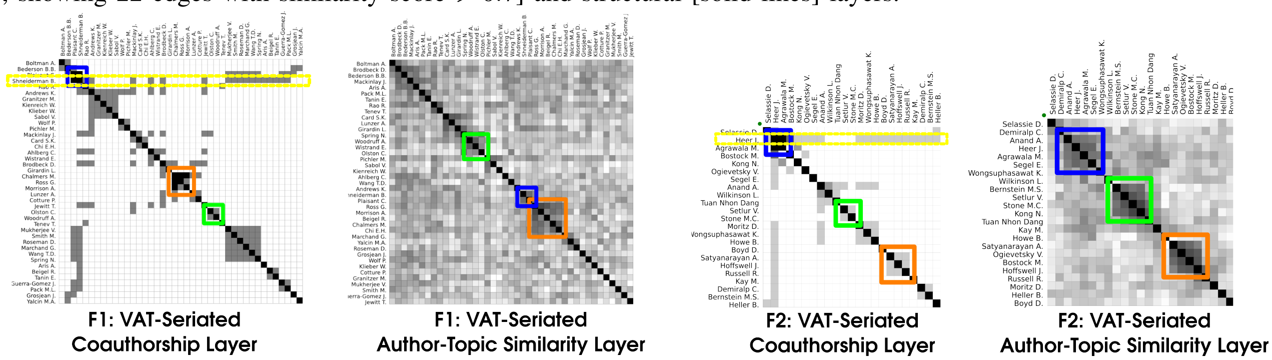


Fig. 3: An aggregated node showing a community in structural layer of IV dataset, after VAT seriation shows clusters recurring in both structural and functional layers [green, blue, orange]. The yellow highlights show central actors.

the user can choose a focus and activate it, and by locking it, the user activates the context. A blue lock icon in the top left corner of the matrix indicates active or locked state, respectively. A focus can be activated by clicking in the region of the AN. When a focus is deactivated, the user can choose another AN as focus. Extending the layout in NodeTrix to render the focus, we additionally render inter-community links from the AN representing the focus. These inter-community links exist in the layer, which is used for visualizing the focus; while we also render (inter-community) links between ANs in the layer used for visualizing the context.

Software Implementation:: NTM has been implemented using Python v2.7 for data preprocessing, Flask framework, and D3.js [9] for visualization. D3.js enables us to perform progressive rendering of sparse links when moving the ANs.

V. Case-Study of a Multiplex Collaboration Network

Our case study, Infovis (IV) co-authorship network [13] during (1995-2015) has 1235 nodes, 2705 edges, 150 communities (detected using Louvain CD). The two layers in IV dataset are co-authorship (structural) and author-topic similarity [22] (functional) graphs. The co-authorship layer (Figure 2) has links between authors if the authors have co-authored, and the edge weight is the number of papers they have co-authored in the topic of Infovis during 1995-2015. The following metadata for each paper is available in the IV dataset: title, authors, keywords, abstract, and references. We have used the metadata to compute the author-topic similarity matrix, which is the adjacency matrix of a similarity graph. Similar to NodeTrix, tasks T2 and T3 can be accomplished from NTM, where the mostly colored row and column (yellow highlights in Figure 3) pertaining to Ben Shneiderman and Jeff Heer, show them to be the central actor in the communities in foci F1 and F2, respectively. Similarly, S. Carpendale, C. North, P. Hanrahan, J. Wood, J. Fekete, J. Dykes, and H. Hauser are central actors in their respective communities/ANs.

NTM helps us find clusters along the diagonal, given by VAT, which recur in multiple layers; e.g. blue, green, and orange highlights in Figure 3 who group together in both the layers. Thus, the staging area (Figure 1(b)) helps in accomplishing task T4. The semantics of such a cluster is that, co-authors in it publish in similar topics, even in papers other than their joint papers. In such clusters in F1 and F2, which also contain the central actors (blue highlights in Figure 3), we observe that the cluster in the structural layer are rendered darker than those in the functional layer, which indicates more accurate similarity scores. On the contrary, the reverse observation in the orange (in F2) and green (in F1 and F2) highlights, where the cluster in the functional layer is darker than its counterpart in the structural layer, indicates erroneous computation of the similarity scores. We have found out that the error in author-topic similarity arises owing to the authors having only one paper in the dataset. Author-topic similarity score is computed using a mixture of distributions associated with the authors in a multi-author paper. A cluster, which is darker in the structural layer than the functional one, implies that the authors have co-authored multiple papers together, owing to which the author-topic similarity scores are more accurate. e.g. {Shneiderman, Plaisant} and {Heer, Agrawala} have authored {8, 4} and {17, 6} independently, and 2 and 5 papers jointly, and thus, have more accurate author-topic similarity scores, 0.57 and 0.60, respectively. Thus, our visualization not only identifies clusters that recur across layers, the aforementioned pattern can help ascertain the accuracy of the results. A corollary to T4 would be to find authors who have not co-authored but have a high author-topic similarity score, which may indicate potential collaboration outside of this network, e.g. {Heer, Stone}[3]. However, these aforementioned patterns are specific to the current scenario of co-authorship and author-topic similarity layers, and should not be generalized. Nonetheless, NTM enables identification of such trends.

NTM is designed to study all aspects of the subnetwork, corresponding to the focus (which is in structural layer), in all functional layers, for visual analytics; without assuming that the focus remains a community across all functional layers. e.g., links in the similarity layer, but between the AN's in the SWN layer, give more information about the overlap of topics the authors work with, thus accomplishing task T5 (Figure 2). Between ANs with Hauser and Shneiderman as central actors, links {Ledermann, Aris} and {Doleisch, Aris} have been observed to exist due to common topics of *plots* and *user interactions*; and {Hauser, Yalcin}, due to the topic of *set visualizations*. Similarly between ANs with Fekete and Shneiderman as central actors, links {Henry, Woodruff} and {Ghoniem, Sabol} have been observed to indicate common topics of*multiple views* and *graph visualization*, respectively.

Work-flow for Community Exploration:: Our work-flow for CD and exploration in a multiplex network, using NTM GUI, is a four-step process (Figure 1). Firstly, we input a multiplex network, $\mathscr{M}$, with N layers, and set the structural layer $\mathscr{E}^0$. In our implementation, we construct the multiplex network using author-topic similarity graph, which is the adjacency graph of a functional layer. Similar to NodeTrix [12], NTM becomes slow for interactive response, when the entire network is loaded. For interactive performance, in our case study, we have used Louvain CD ($\mathscr{O}(|\mathscr{V}(\mathscr{M})|\log(|\mathscr{V}(\mathscr{M})|))$ complexity) to identify communities on the structural layer of the entire network, to find logical subnetworks of size upto 250 nodes, to be loaded on NTM. Here, we have used the vertex set of three largest communities in the network as our subnetwork of interest. This step will, however, not be required once NTM is scaled to handle loading of the entire network. Secondly, Louvain CD is performed on the structural layer of the subnetwork, which is loaded on NTM, as a preprocessing step. In our specific case, performing Louvain CD on the entire network and on the subnetwork yield different results; hence, we repeat running the algorithm on the subnetwork after it is loaded. Thirdly, the user can interact with the tool, and pick an AN as a focus. Fourthly, the user can build multiple HDMs, and perform automatic seriation on the AN, using VAT, to visualize possible clusters in each of the layers. For further analysis, different images of the focus are saved and loaded in the staging area.

Expert User Evaluation:: We have performed an expert user evaluation of the tool, which is built using NTM as a framework and is available at http://nmultiplex.au-syd.mybluemix.net/ . The expert, who is a network science researcher, analyzed the usefulness and usability of the tool. The expert mentioned that the use of focus+context visualization helps in focused analysis of communities and hence, the HDM is useful. We have presented the visualizations of the HDM in an existing tool, Gephi (Figure 4), and NTM (Figures 1 and 2), to the expert. The expert mentioned that the visualizations are better readable on NTM than on Gephi. The expert commented that the HDM and the tool are useful for finding relevant nested communities, which gives a *mesoscopic* network analysis. The ability to switch across different layers allows the user to get an overview of the dynamics occurring in each layer. While the tool does not automate community analysis across the layers, the expert was able to study each focus in detail using the tool. However, the tool is limited in answering specific questions within foci or communities alone, and in its current state, the tool cannot perform a generic analysis of all communities. It also cannot give comparisons of the "strength" of communities across layers. Nevertheless, overall evaluation has been encouraging.
Usability Evaluation:: The expert commented that the tool is predominantly easy to use, with the help of the interactive tutorial. The interactivity is responsive, especially due to updates using progressive rendering. The expert liked the color combinations for improving the visual experience. At the same time, the expert pointed out the limitations in the usability of the current version, such as overloading of features on the right mouse button and non-intuitive user interaction for panning in the main view. Currently, the right mouse button is used for selecting focus, popping up the browser menu, and dragging the focus; the

[3]Lin, Sharon, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. "Selecting Semantically]Resonant Colors for Data Visualization." In Computer Graphics Forum, vol. 32, no. 3pt4, pp. 401-410. Blackwell Publishing Ltd, 2013.

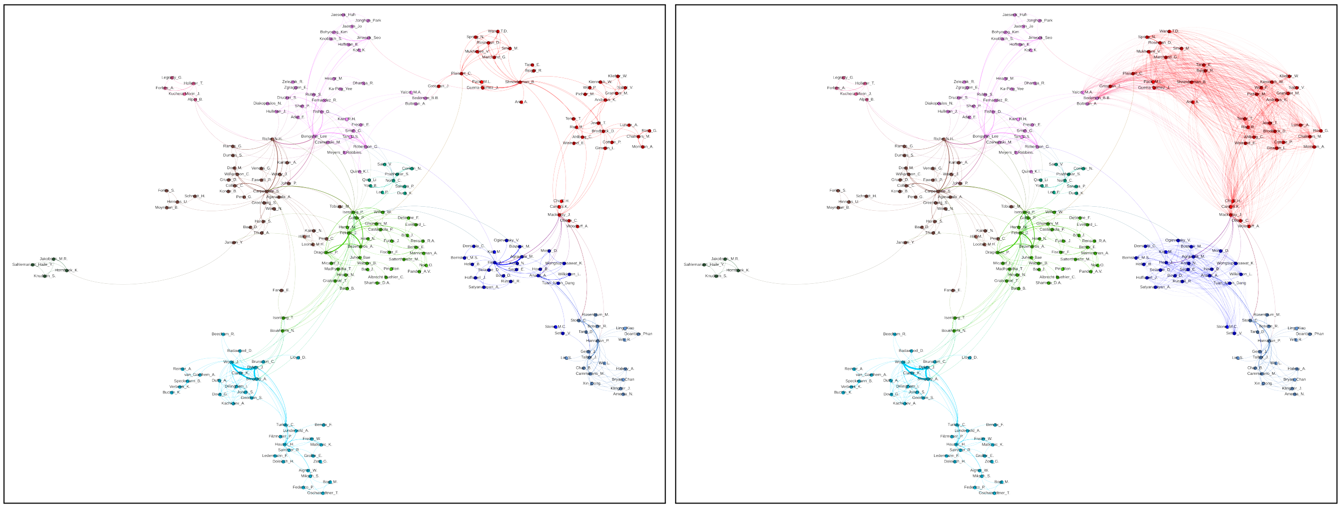


Fig. 4: An equivalent of graph layout in Gephi of the subnetwork of interest, showing the communities in the structural layer, detected using Louvain CD in different colors. Foci F1 and F2 in red and blue, in the structural layer in left, and in similarity layer in the right. The latter shows the node-link diagram of the HDM.

scroll wheel is used for zooming in and out; and dragging the left and right mouse buttons has been used for panning. The limitations can be alleviated with UI re-design of the tool.

VI. Conclusions

We have proposed and implemented a visual analytic framework, NTM, for probing a subnetwork of interest, chosen as a focus, in a multiplex SWN. We have used a focus+context paradigm, our proposed HDM, visual analytic workflow and seriation for clustering. We have constructed a multiplex network from a co-authorship network (structural layer) by computing author-topic similarity graph as the functional layer. However, there are few limitations in our current approach. In this work, we have focused on multiplex SWNs, owing to which the network topology of the structural layer is restrictive. At the same time, in order to extend this work to different kinds of multiplex networks, without none of the layers exhibiting the small world property, we need to consider an appropriate visual representation of the concerned network topology. NTM, being an extension of NodeTrix, is effective as a hybrid visualization of node-link diagrams and matrix visualization, as the "globally sparse" property of the SWNs reduces clutter and occlusion in the visualization. If the intercommunity links were not to be as sparse as seen in the SWN topology, then the hybrid visualization gets very cluttered. We are currently working on improving scalability in using multiplex networks with more than two layers. We are also investigating other graph layouts, without a bias on SWN layer.

VII. Acknowledgements

This work has been partially supported by funding from NRDMS, Department of Science & Technology, Government of India; RSA division of EMC^2 India; and INCOIS, Ministry of Earth Sciences, Government of India. The authors are grateful to Dr. T. K. Srikanth for his valuable contributions in improving the tool, and to the anonymous reviewers for comments in improving the paper.

References

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[2] Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy, et al. Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362, 2009.

[3] Federico Battiston, Vincenzo Nicosia, and Vito Latora. Structural measures for multiplex networks. *Physical Review E*, 89(3):032804, 2014.

[4] Michael Behrisch, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, and Jean-Daniel Fekete. Matrix reordering methods for table and network visualization. In *Computer Graphics Forum*, volume 35, page 24, 2016.

[5] Laura Bennett, Aristotelis Kittas, Gareth Muirhead, Lazaros G Papageorgiou, and Sophia Tsoka. Detection of composite communities in multiplex biological networks. *Scientific reports*, 5, 2015.

[6] James C Bezdek, Richard J Hathaway, and Jacalyn M Huband. Visual assessment of clustering tendency for rectangular dissimilarity matrices. *Fuzzy Systems, IEEE Transactions on*, 15(5):890–903, 2007.

[7] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[8] Stefano Boccaletti, Ginestra Bianconi, Regino Criado, Charo I Del Genio, Jesus Gómez-Gardeñes, Miguel Romance, Irene Sendiña-Nadal, Zhen Wang, and Massimiliano Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.

[9] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.

[10] Manlio De Domenico, Andrea Lancichinetti, Alex Arenas, and Martin Rosvall. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X*, 5(1):011027, 2015.

[11] Mohammad Ghoniem, Jean-Daniel Fekete, and Philippe Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 17–24. Ieee, 2004.

[12] Nathalie Henry, Jean-Daniel Fekete, and Michael J McGuffin. Nodetrix: a hybrid visualization of social networks. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1302–1309, 2007.

[13] Petra Isenberg, Florian Heimerl, Steffen Koch, Tobias Isenberg, Panpan Xu, Chad Stolper, Michael Sedlmair, Jian Chen, Torsten Möller, and John Stasko. Visualization publication dataset. Dataset: http://vispubdata.org/, 2015. Published Jun. 2015.

[14] Ilir Jusufi, Andreas Kerren, and Bjorn Zimmer. Multivariate Network Exploration with JauntyNets. In *Proceedings of the 17th International Conference on Information Visualisation (IV'13)*, pages 19–27. IEEE Computer Society Press, 2013.

[15] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[16] Innar Liiv. Seriation and matrix reordering methods: An historical overview. *Statistical analysis and data mining*, 3(2):70–91, 2010.

[17] Rafael Messias Martins, Gabriel Faria Andery, Henry Heberle, Fernando Vieira Paulovich, Alneu de Andrade Lopes, Helio Pedrini, and Rosane Minghim. Multidimensional projections for visual analysis of social networks. *Journal of Computer Science and Technology*, 27(4):791–810, 2012.

[18] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *science*, 328(5980):876–878, 2010.

[19] Hae-Jeong Park and Karl Friston. Structural and functional brain networks: from connections to cognition. *Science*, 342(6158):1238411, 2013.

[20] Saima Parveen and Jaya Sreevalsan-Nair. Visualization of small world networks using similarity matrices. In *Big Data Analytics*, pages 151–170. Springer, 2013.

[21] Benjamin Renoust, Guy Melançon, and Tamara Munzner. Detangler: Visual analytics for multiplex networks. In *Computer Graphics Forum*, volume 34, pages 321–330. Wiley Online Library, 2015.

[22] Michal Rosen-Zvi, Chaitanya Chemudugunta, Thomas Griffiths, Padhraic Smyth, and Mark Steyvers. Learning author-topic models from text corpora. *ACM Transactions on Information Systems (TOIS)*, 28(1):4, 2010.

[23] Luca Rossi and Matteo Magnani. Towards effective visual analytics on multiplex and multilayer networks. *Chaos, Solitons & Fractals*, 72:68–76, 2015.

[24] Sébastien Rufiange, Michael J McGuffin, and Christopher P Fuhrman. Treematrix: A hybrid visualization of compound graphs. In *Computer Graphics Forum*, volume 31, pages 89–101. Wiley Online Library, 2012.

[25] Lei Shi, Nan Cao, Shixia Liu, Weihong Qian, Li Tan, Guodong Wang, Jimeng Sun, and Ching-Yung Lin. Himap: Adaptive visualization of large-scale online social networks. In *Visualization Symposium, 2009. PacificVis' 09. IEEE Pacific*, pages 41–48. IEEE, 2009.

[26] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

[27] Robert Tarjan. Depth-first search and linear graph algorithms. *SIAM journal on computing*, 1(2):146–160, 1972.

[28] Stef van den Elzen and Jarke J van Wijk. Multivariate network exploration and presentation: From detail to overview via selections and aggregations. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2310–2319, 2014.

[29] Corinna Vehlow, Fabian Beck, and Daniel Weiskopf. The state of the art in visualizing group structures in graphs. In *Eurographics Conference on Visualization (EuroVis)-STARs*, pages 21–40, 2015.

[30] Corinna Vehlow, Thomas Reinhardt, and Daniel Weiskopf. Visualizing fuzzy overlapping communities in networks. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2486–2495, 2013.

[31] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684):440–442, 1998.