2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Local Community Detection in Multilayer Networks

R. Interdonato, A. Tagarelli
DIMES, Università della Calabria, Italy

D. Ienco[(1)], A. Sallaberry[(2)], P. Poncelet[(3)]
[(1)]IRSTEA, UMR TETIS,
[(2)]LIRMM - Université Paul Valéry,
[(3)]LIRMM - Université de Montpellier, France

***Abstract*—The problem of local community detection refers to the identification of a community starting from a query node and using limited information about the network structure. Existing methods for solving this problem however are not designed to deal with multilayer network models, which are becoming pervasive in many fields of science. In this work, we present the first method for local community detection in multilayer networks. Our method exploits both internal and external connectivity of the nodes in the community being constructed for a given seed, while accounting for different layer-specific topological information. Evaluation of the proposed method has been conducted on real-world multilayer networks.**

I. Introduction

Local community detection [1], [2] is the problem of identifying a community structure which is centered on one or few seed users, given limited information about the network. Despite this problem has gained increasing interest in the last few years, it has been mainly investigated by focusing on networks that are built on a single node-relation type or context. Note however that individuals often have multiple accounts across different social networks, as well as relations of different types can be available for the same population of a network [3] (e.g., followship, like/comment interactions, working relationship, lunch relationship). These scenarios can effectively be represented using a *multilayer network* model [5].

In this work we propose a method for the novel problem of local community detection in a multilayer network (ML-LCD), following an unsupervised paradigm that exploits layer-specific topological information. We evaluated our method on real-world multilayer networks; due to space limits, we report here part of an analysis of structural characteristics of the extracted local communities. To the best of our knowledge, we are the first to bring the local community detection problem into the context of multilayer networks, since all previous works have addressed the community detection task on multilayer networks from a global point of view [4], [5].

II. Multilayer Local Community Detection

Multilayer network model. We refer to the general multilayer network model described in [5]. We are given a set of layers $\mathcal{L}$ and a set of users $\mathcal{V}$. We denote with $G_{\mathcal{L}} = (V_{\mathcal{L}}, E_{\mathcal{L}}, \mathcal{V}, \mathcal{L})$ the multilayer graph such that $V_{\mathcal{L}}$ is a set of pairs $v \in \mathcal{V}, L \in \mathcal{L}$, an $E_{\mathcal{L}} \subseteq V_{\mathcal{L}} \times V_{\mathcal{L}}$ is the set of undirected edges. Note that we do not require all nodes (elements of $\mathcal{V}$) participate to all layers, however each node appears in at least one layer. Moreover, the only inter-layer edges are those for which the two nodes represent the same entity (i.e., element of $\mathcal{V}$) in different layers.

Method. Local community detection approaches generally implement some strategy that at each step considers a node from one of three sets, namely: the community under construction (initialized with the seed node), the "shell" of nodes that are neighbors of nodes in the community but do not belong to the community, and the unexplored portion of the network. A key aspect is hence how to select the *best* node in the shell to add to the community to be identified. Most algorithms account for the relative ratio of internal edges (i.e., links between nodes in the community) and external edges (i.e., links between nodes in the community and nodes outside the community), therefore they penalize candidates in proportion to the amount of links to non-community nodes [2].

Our approach follows the above general strategy. Nevertheless, to account for the multiplicity of layers, we introduce a multilayer local community function that relies on a notion of similarity of nodes. In this regard, two major issues are how to choose the analytical form of the similarity function, and how to deal with the different, layer-specific connections that any two nodes might have in the multilayer graph. We address the first issue in an unsupervised fashion, by resorting to any similarity measure that can express the topological affinity of two nodes in a graph. Concerning the second issue, one straightforward solution is to determine the similarity between any two nodes focusing on each layer at a time. The above points are formally captured by the following definitions.

Given $G_{\mathcal{L}} = (V_{\mathcal{L}}, E_{\mathcal{L}}, \mathcal{V}, \mathcal{L})$, and any seed node v_0, we will use symbol C to denote the subgraph corresponding to the local community built around node v_0. We denote with $S = \{v \in \mathcal{V} \setminus C | (v, u) \in E_{\mathcal{L}}, u \in C\}$ the *shell* set of nodes outside C, and with $B = \{v \in C | \exists (u, v) \in E_{\mathcal{L}}, u \in S\}$ the *boundary* set of nodes in C. Moreover, we denote with E^C the set of edges between nodes that belong to C and with E_i^C the subset of edges that correspond to a given layer L_i. Analogously, E^B refers to the set of edges between nodes in B and nodes in S, and E_i^B to its subset corresponding to L_i.

Given a community C, we define the *similarity-based local community measure* $LC(C)$ as the ratio between two terms. The first term expresses an *internal community relation* we

IEEE/ACM ASONAM 2016, August 18-21, 2016, San Francisco, CA, USA
978-1-5090-2846-7/16/$31.00 © 2016 IEEE

define as:

$$LC^{int}(C) = \frac{1}{|C|} \sum_{v \in C} \sum_{L_i \in \mathcal{L}} \sum_{\substack{(u,v) \in E_i^C \\ \wedge u \in C}} sim_i(u,v) \qquad (1)$$

and the second term expresses an *external community relation*:

$$LC^{ext}(C) = \frac{1}{|B|} \sum_{v \in B} \sum_{L_i \in \mathcal{L}} \sum_{\substack{(u,v) \in E_i^B \\ \wedge u \in S}} sim_i(u,v) \qquad (2)$$

In the above equations, function $sim_i(u,v)$ computes the similarity between any two nodes u, v contextually to layer L_i. In this work, we define it in terms of Jaccard similarity, i.e., $sim_i(u,v) = \frac{|N_i(u) \cap N_i(v)|}{|N_i(u) \cup N_i(v)|}$, where $N_i(u)$ denotes the set of neighbors of node u in layer L_i.

The proposed **M***ulti***L***ayer* **L***ocal* **C***ommunity* **D***etection* (ML-LCD) algorithm takes as input the multilayer graph $G_{\mathcal{L}}$ and a seed node v_0, and computes the local community C associated to v_0 by performing an iterative search that seeks to maximize the ratio of $LC^{int}(C)$ to $LC^{ext}(C)$.

Initially, the boundary set B and the community C are initialized with the starting seed, while the shell set S is initialized with the neighborhood set of v_0 considering all the layers in $\mathcal{L}$. Afterwards, the algorithm computes the initial value of $LC(C)$ and starts expanding the node set in C: it evaluates all the nodes v belonging to the current shell set S, then selects the vertex v^* that maximizes the value of $LC(C)$. The algorithm checks if *(i)* v^* actually increases the quality of C (i.e., $LC(C \cup \{v^*\}) > LC(C)$) and *(ii)* v^* helps to strength the internal connectivity of the community (i.e., $LC^{int}(C \cup \{v^*\}) > LC^{int}(C)$). If both conditions are satisfied, node v^* is added to C and the shell set is updated accordingly, otherwise node v^* is removed from S as it cannot lead to an increase in the value of $LC(C)$. In any case, the boundary set B and $LC(C)$ are updated. The algorithm terminates when no further improvement in $LC(C)$ is possible.

III. Experimental Results

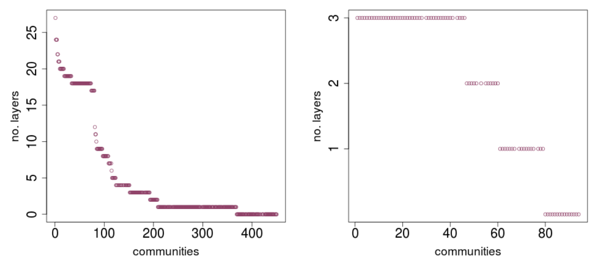
We used three real-world multilayer network datasets, namely *Airlines* (417 nodes corresponding to airport locations, 3588 edges, 37 layers corresponding to airline companies) [6], *AUCS* (61 employees as nodes, 620 edges, 5 acquaintance relations as layers) [3], and *RealityMining* (88 users as nodes, 355 edges, 3 media types employed to communicate as layers) [7]. In all datasets, node relations are symmetric.

We analyzed structural characteristics of the local communities extracted by ML-LCD for each node, over all networks. Largest local communities were observed for *Airlines* (mean 11.48 ± 15.04), while medium size communities (7.90 ± 2.74) were discovered for *AUCS* and relatively small communities (3.37 ± 1.77) for *RealityMining*.

Table I shows the per-layer *average path length* and *clustering coefficient* of the identified communities. For each of the datasets and measures, we report mean and standard deviation over the layers, of the mean and maximum values found over all communities. Maximum values of average path length are quite similar for all datasets, while average values are relatively small (i.e., below 1.0) for *Airlines* and *RealityMining*, and slightly higher for *AUCS*. By coupling these results with clustering coefficient values, roughly small-world communities are observed on *AUCS* and *RealityMining*.

TABLE I
Per-layer average path length and clustering coefficient

Dataset	average path length		clustering coefficient	
	Avg.	Max	Avg.	Max
Airlines	0.222 ± 0.168	1.872 ± 0.397	0.022 ± 0.027	0.508 ± 0.398
AUCS	1.182 ± 0.237	1.880 ± 0.364	0.533 ± 0.238	0.938 ± 0.087
RealityMining	0.778 ± 0.111	1.833 ± 0.153	0.295 ± 0.126	1.000 ± 0.000



(a) *Airlines* (b) *RealityMining*

Fig. 1. Distribution of number of layers over communities.

Considering the amount of layers covered by each particular community, Figure 1 reports results for *Airlines* and *RealityMining*. (Communities are sorted by decreasing number of layers.) On the former, we observe that ML-LCD produces about 25% of the communities covering 50-70% of layers. On the latter, ML-LCD is able to produce most of the communities that cover all layers, following a stairs-like behavior for communities with a lower number of layers.

IV. Conclusion

We addressed the novel problem of local community detection in multilayer networks, and presented the first method to solve it. This employs a greedy heuristic that considers both internal and external connectivity, following an unsupervised paradigm that exploits layer-specific topological information. Evaluation was conducted on real-world multilayer networks.

References

[1] A. Clauset, "Finding local community structure in networks," *Phys. Rev. E*, vol. 72, no. 2, p. 026132, 2005.

[2] J. Chen, O. R. Zaïane, and R. Goebel, "Local community identification in social networks," in *Proc. ASONAM*, 2009, pp. 237–242.

[3] M. E. Dickison, M. Magnani, and L. Rossi, *Multilayer social networks*. Cambridge University Press, 2016.

[4] C. W. Loe and H. J. Jensen, "Comparison of communities detection algorithms for multiplex," *Physica A*, vol. 431, pp. 29–45, 2015.

[5] M. Kivela, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Mutilayer networks," *Journal of Complex Networks*, vol. 2, no. 3, pp. 203–271, 2014.

[6] A. Cardillo, J. Gomez-Gardenes, M. Zanin, M. Romance, D. Papo, F. del Pozo, and S. Boccaletti, "Emergence of network features from multiplexity," *Scientific Reports*, vol. 3:1344, 2013.

[7] J. Kim and J. Lee, "Community detection in multi-layer graphs: A survey," *SIGMOD Record*, vol. 44, no. 3, pp. 37–48, 2015.