Local community detection in multilayer networks

Roberto Interdonato, Andrea Tagarelli, Dino Ienco, Arnaud Sallaberry, Pascal Poncelet

Local Community Detection in Multilayer Networks

R. Interdonato [(1)], A. Tagarelli [(1)], D. Ienco [(2)], A. Sallaberry[(3)], P. Poncelet[(4)]

[(1)]DIMES, Università della Calabria, Italy [(2)]IRSTEA, UMR TETIS,
[(3)]LIRMM - Université Paul Valéry, [(4)]LIRMM - Université de Montpellier, France

Introduction

Local community detection [2, 1] is the problem of identifying a community structure which is centered on one or few seed users. In contrast to traditional community detection, which requires knowledge on the whole network structure, methods for local community detection can handle limited information about the network.

Despite the local community detection 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). There is indeed emergence for developing methods that solve the local community detection over such scenarios, which can be conveniently represented using a *multilayer network* model [4]. Figure 1 illustrates an example of multilayer local community, centered on a given seed node, that is identified over a multilayer network.

Contributions

- We propose the first method for local community detection in multilayer networks (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 [5, 4].

MultiLayer Local Community Detection

Multilayer network model. 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 $v \in \mathcal{V}, L \in \mathcal{L}$, and $E_{\mathcal{L}} \subseteq V_{\mathcal{L}} \times V_{\mathcal{L}}$ is the set of undirected edges. We do not require all nodes (elements of $\mathcal{V}$) participate to all layers, however each node appears in at least one layer. Only inter-layer edges correspond to links between nodes representing the same entity (i.e., element of $\mathcal{V}$) in different layers.

We denote with C the subgraph corresponding to the local community built around the seed node v_0. $S = \{v \in \mathcal{V} \setminus C | (v, u) \in E_{\mathcal{L}}, u \in C\}$ is the *shell* set of nodes outside C, and $B = \{v \in C | \exists (u, v) \in E_{\mathcal{L}}, u \in S\}$ is the *boundary* set of nodes in C. E_i^C is the subset of edges among nodes in C that corresponds to a given layer L_i. E_i^B refers to the subset of edges between nodes in B and nodes in S corresponding to L_i.

Algorithm. The proposed **M**ulti**L**ayer **L**ocal **C**ommunity **D**etection (ML-LCD) algorithm takes as input $G_{\mathcal{L}}$ and v_0, and computes the local community C associated to v_0 by performing an iterative search that seeks to maximize the the *similarity-based local community measure* $LC(C)$. The latter is defined as the ratio of $LC^{int}(C)$, i.e., *internal community relation*, to $LC^{ext}(C)$, i.e., *external community relation*, which are in turn defined as:

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

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

where $sim_i(u, v)$ computes the similarity between any two nodes u, v contextually to layer L_i (e.g., Jaccard similarity between sets of neighbors).

Example of Multilayer Local Community

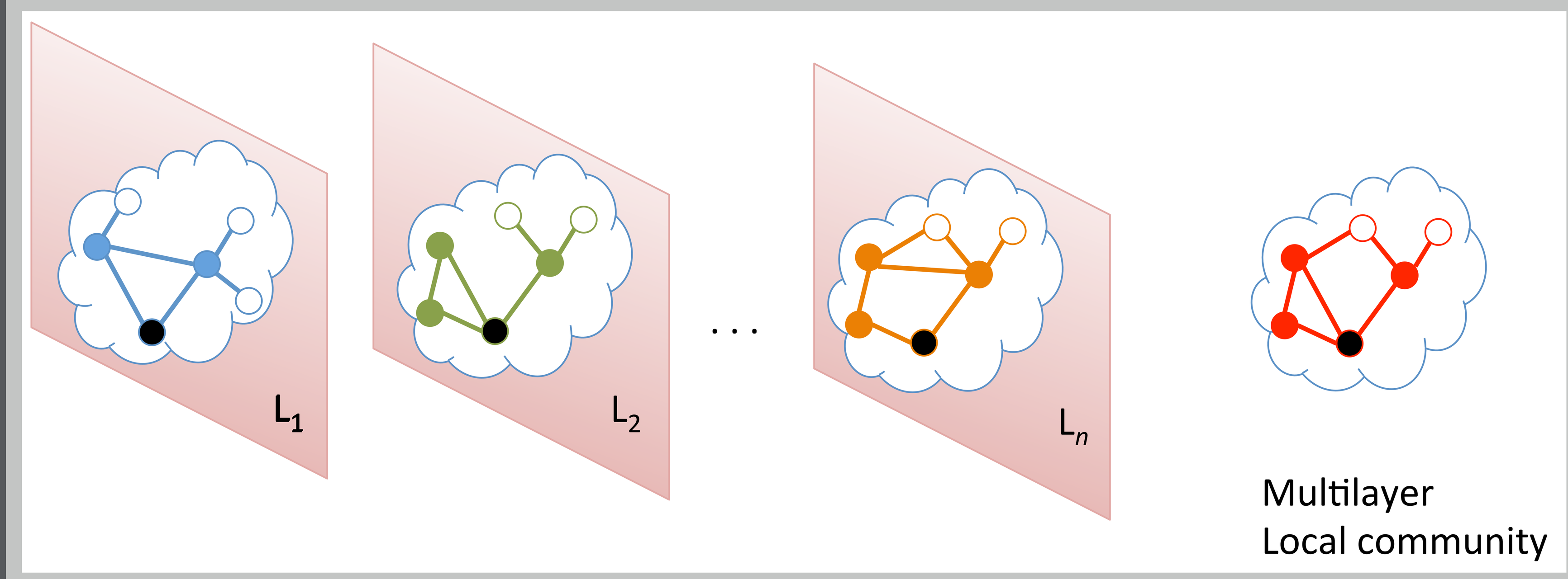

Figure 1: A local community identified over an example n-layer network (on the right) and the corresponding projections over each of the layers (on the left). Seed node, within-community nodes and shell nodes are denoted with black filled, colored and empty circles, respectively. Clouds delimit the boundaries between the shell set and the unknown part of the network graph.

Experimental Results

Table 1: Main characteristics of the multilayer network datasets

Dataset	# Nodes	# Edges	#Layers	Density	A_{deg}	A_{layer}
Airlines [6]	417	3588	37	0.056	17.21	4.88
AUCS [3]	61	620	5	0.114	20.33	3.67
RealityMining [7]	88	355	3	0.047	8.07	2.42

Table 2: Mean/std size, average path length and clustering coefficient, averaged over the communities extracted by ML-LCD

Dataset	mean community size	average path length		clustering coefficient	
	mean ± sd	Avg.	Max	Avg.	Max
Airlines	11.48 ± 15.04	0.222 ± 0.168	1.872 ± 0.397	0.022 ± 0.027	0.508 ± 0.398
AUCS	7.90 ± 2.74	1.182 ± 0.237	1.880 ± 0.364	0.533 ± 0.238	0.938 ± 0.087
RealityMining	3.37 ± 1.77	0.778 ± 0.111	1.833 ± 0.153	0.295 ± 0.126	1.000 ± 0.000

- Largest local communities were observed for *Airlines*, while medium size communities were discovered for *AUCS* and relatively small communities for *RealityMining*.
- Maximum values of average path length are quite similar for all datasets, while average values are relatively smaller for *Airlines* and *RealityMining*. Roughly small-world communities are observed on *AUCS* and *RealityMining*.
- Considering the amount of layers covered by each particular community, on *Airlines* about 25% of the communities cover 50-70% of layers, while on *RealityMining* most of the communities cover all layers.

References

[1] J. Chen, O. R. Zaïane, and R. Goebel. “Local community identification in social networks”. In *Proc. ASONAM*, pages 237–242, 2009.

[2] A. Clauset. “Finding local community structure in networks”. *Phys. Rev. E*, 72(2):026132, 2005.

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

[4] M. Kivela, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. “Mutilayer networks”. *Journal of Complex Networks*, 2(3):203–271, 2014.

[5] C. W. Loe and H. J. Jensen. “Comparison of communities detection algorithms for multiplex”. *Physica A*, 431:29–45, 2015.

[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*, 3:1344, 2013.

[7] J. Kim and J.-G. Lee. “Community detection in multi-layer graphs: A survey”. *SIGMOD Record*, 44(3):37–48, 2015.

Contact Information

- Web: http://uweb.deis.unical.it/tagarelli/
- Email: andrea.tagarelli@unical.it, dino.ienco@irstea.fr