Searching communities' in badly conditioned graphs (and 4)

J.M. Montañana*, A. Hervás†,1, and S. Morillas‡

(*) inpeek GmbH, Darmstadt, Germany,
Rheinstraße 40-42, 64283 Darmstadt, Germany.

(†) I.U. de Matemática Multidisciplinar, Universitat Politècnica de València, Spain,
Camí de Vera s/n, València, Spain.

(‡) IUMPA, Universitat Politècnica de València, València, Spain,
Camí de Vera s/n, València, Spain.

1 Introduction

In the study of complex networks, we refer to communities as subgraphs in which their nodes are strongly related to each other and little to the rest of the nodes of the network. These vertex groupings, with common characteristics, play a defining role in the network s tructure. The detection of communities allows us to identify internal structures in a graph that divide the network into groups of vertices with dense internal relationships and dispersed between different communities. Communities play an important role in problems such as student movements, food chain, airport traffic, de termination of pu blic tr ansport ro utes an d ot her ty pes of mo bility, pr opagation of false news, metabolic reactions, brain networks, etc. [1, 7, 16, 17]

The detection of communities has received wide attention in recent years, and multiple methods have been developed to obtain them, especially in the case of untreated graphs. Many algorithms use modularity as an objective function but depending on the criteria or approach of the procedure, different c ommunities a re o btained w ith s imilar m odularity. [4, 7]. C lassical m ethods u se graph partitioning, statistical inference, optimization methods, clustering, center detection, etc. [5, 6, 12, 17]. Recently new methods have been introduced that bring new ideas, from the use of convolution and calculation of borders, genetic algorithms, and deep learning. [8, 11, 14, 15]

In general, the methods are presented to the extent that they are adapted to the characteristics of the problem to be solved, in any case there is a wide collection of methods that work satisfactorily in most cases, particularly for non-directed graphs. Unfortunately, in the case of directed graphs, the algorithms give less satisfactory results.

2 Background

There are several techniques and algorithms that have been developed for the detection and study of communities within graphs. In the Figure 1 is shown the results of applying some of the most popular algorithms on the Zachary's karate club graph, which is a not directed graph. The figure also shows that depending on the algorithm it is obtained different communities.

[1]ahervas@imm.upv.es

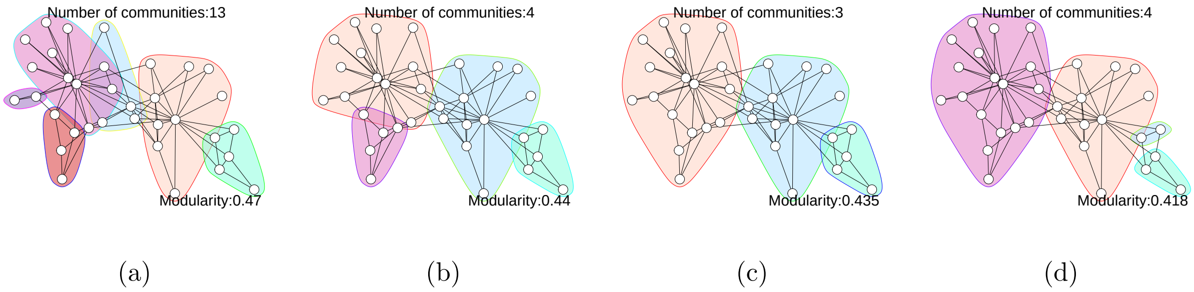


Figure 1: Results obtained when applying (a) Girvan-Newman, (b) Walktrap, (c) Fast-greedy, and (d) Label-propagation community detection algorithms

However, in the case of directed graphs, weakly connected and poorly conditioned, the algorithms for finding communities do not work satisfactorily.

The Figure 2 shows the results when applying these detection algorithms to a directed graphs, weakly connected and poorly conditioned generated with a synthetic generator [9]. The results show a great variation and differ considerably from the expected result.

The generated synthetic directed graph has 8 communities, 37 vertices, 112 edges. The results obtained when applying standard community detection methods are very different, as shown in Figures 2.a-2.d. Figure 2.a shows a result with low modularity and fewer communities. While the results in the other cases have high modularity but the communities obtained differ from those that really exist in the graph.

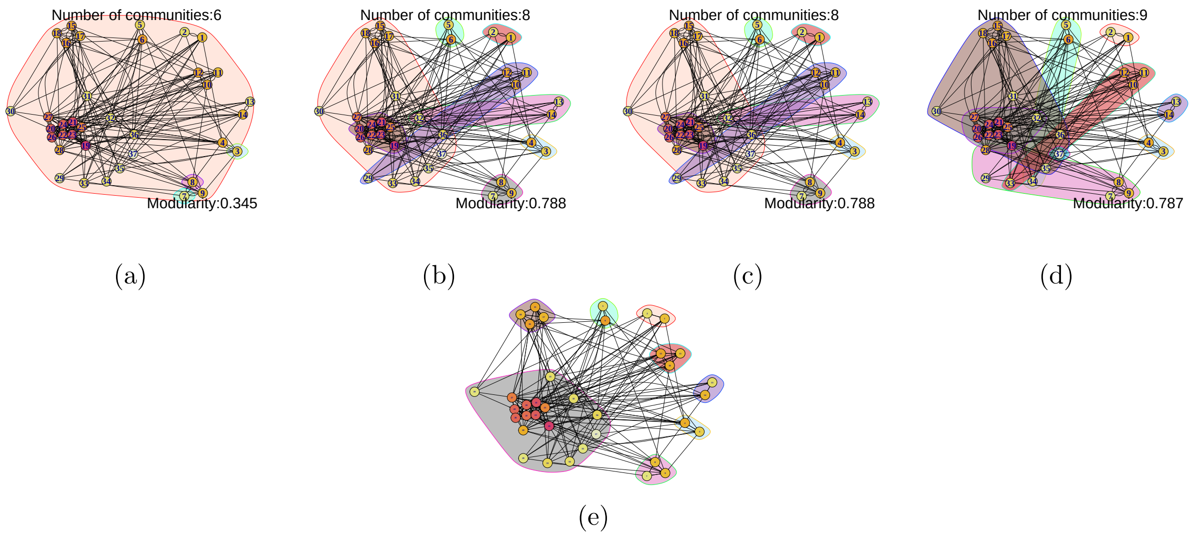


Figure 2: Results obtained when applying (a) Girvan-Newman, (b) Walktrap, (c) Fast-greedy, and (d) Label-propagation community detection algorithms. And (e) the expected detection of communities

We considered also the pruning method from [11]. However, that method did not work in the case of directed graphs, weakly connected and poorly conditioned.

3 Method

We take as inspiration the metrics used for detecting edges in images, to define a metric that allows to define a quantitative value for each edge in the graph, that we can use to define the relevance of the edge among its neighbor edges. This metric allows to define an objective and systematic way to compare the edges, and in our particular case, to identify edges between communities. This case is of particular interest, because such kind of edges can be present on the experimental results due to noise or errors during the collection of data. Having a systematic way for their identification and pruning allows to the standard algorithms be able to provide more representative identification of communities with the cases of study.

In this paper, we target to define a Metric for evaluating the relevance or significance of the edges, and then prune only the less relevant ones that induce a incorrect identification of communities, which correspond to edges on the boundary or transition between communities. Then, the existing community detection algorithms can be applied on those graphs after being pruned.

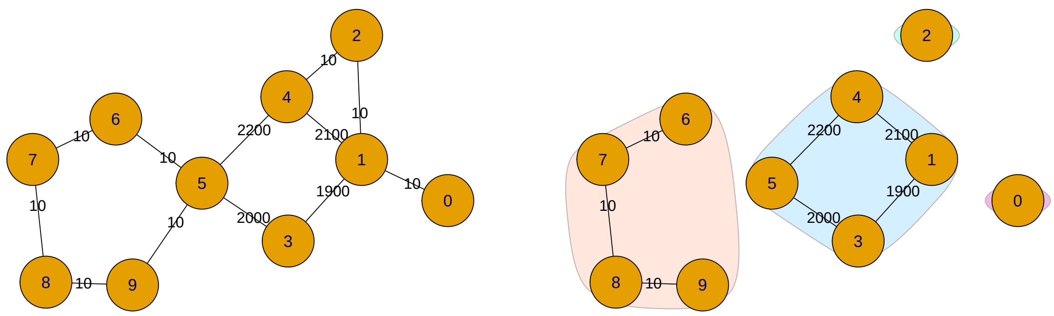
We define a metric at equation 1 that resembles a metric used for the detection of border of areas in a image by evaluating a function among a pixel and its neighboring pixels.

$$M(s,t) = \frac{1}{1 + \left|(n-1)v'_s - \sum_{j=1}^{j<=n} v'_j\right|} \quad \text{where } j \text{ are the adjacent elements to } s \text{ at the linear graph} \tag{1}$$

We define the pruning condition of an edge when its corresponding value contained in the module value in the metric is negative.

4 Results

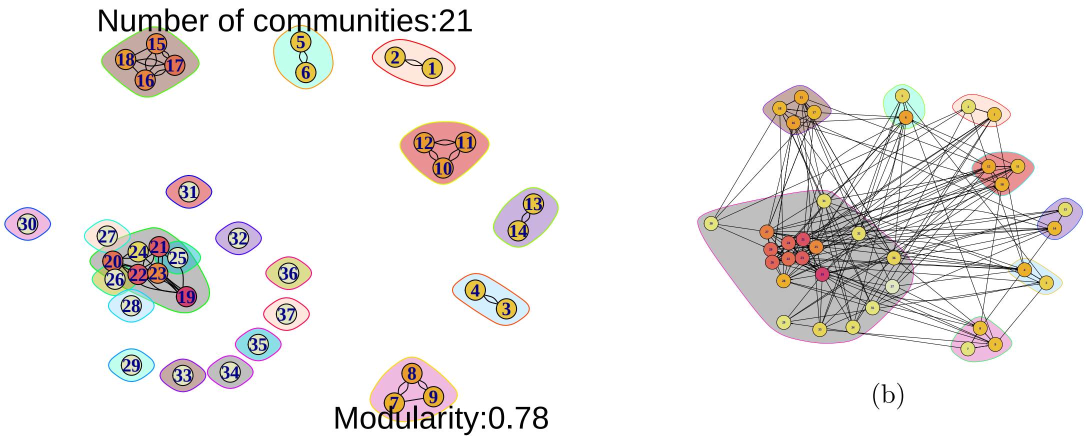
The results of applying the pruning method works satisfactorily when applying to the example provided at [11]. As shown in the next figure.



(a) (b)

Figure 3: (a) Original graph provided at [11], and (b) pruned graph with the proposed method in this paper

Moreover, our method also is able to remove the links between the communities in the case of directed graphs, weakly connected and poorly conditioned. As an example, the next figure shows the result of pruning the graph considered in the previous section on which none of the evaluated community detection algorithm was able to provide the expected detection.



(b)

(a)

Figure 4: Results obtained when applying (a) Girvan-Newman community detection algorithm on the pruned graph, and (b) the expected detection of communities

5 Conclusions

In this paper, a novel approach is proposed, leveraged on a **fuzzy metric** for pruning graphs, including directed weakly connected and poorly conditioned graphs. The results obtained improve the results obtained with other methods. As the next steps, we are going to refine and validate our metric applying it to larger complex networks obtained from real scenarios as well as synthetic ones.

References

[1] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. U. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5), 175–308, 2006.

[2] Deng, Y., Kenney, C.,Moore, M. S., and Manjunath, B. S. Peer group filtering and perceptutal color image quantization, in Proc. IEEE Int. Symp. Circuits Systems, 1999, vol. 4, pp. 21-24. 1999.

[3] Feng S, Hu B, Nie C, Shen X. Empirical Study on a Directed and Weighted Bus Transport Network in China. Physica A: Statistical Mechanics and its Applications 441(C):85-92. 2016.

[4] Girvan, M.; Newman, M. E. J. Community structure in social and biological networks". Proceedings of the National Academy of Sciences. 99 (12): 7821-7826. 5. 2002.

[5] Hervás, A., Soriano, P. P., Capilla, R., and Peinado, J. An algorithm to obtain directional communities in a directed graph. In L. Jodar Ed. Proceedings of the Mathematical Modelling in Engineering & Human Behaviour, 347-352. SPUPV, Valencia, Spain. 2016.

[6] Hervás, A., Soriano-Jiménez, P. P., Jiménez, A., Peinado, J., Capilla, R., and Montañana, J. M. A Case Study of Directional Communities in a Directd Graph: The Accessing Procedure to the Spanish Public University System. Modeling Human Behavior, 9. Nova Science Publishers, Inc. Hauppauge, NY. USA. 2015.

[7] Lancichinetti A, and Fortunato S. Community detection algorithms: A comparative analysis. Phys. Rev. E 80(5):056117. 2009.

[8] Liu, H., Wei, J., and Xu, T. Community detection based on community perspective and graph convolutional network. Expert Systems with Applications, 231, 120748. 2023.

[9] Montañana, J. M., Hervás, A., and Soriano Jiménez, P. P. A Proposal for a Benchmark Generator of Weakly Connected Directed Graphs. Open Journal of Modelling and Simulation, 8(1), 18-34. 2020.

[10] Morillas, S., Gregori, V., and Hervás, A. Fuzzy peer groups for reducing mixed Gaussian-impulse noise from color images. IEEE Transactions on Image Processing, 18(7), 1452-1466. 2009.

[11] Muñoz,H., Vicente,E., Gonzalez,I., Mateos, A., and A. Jimenez-Martin. ConvGraph: Community Detection of Homogeneous Relationships in Weighted Graphs, Mathematics, vol. 9, no 4, pag. 367. 2021.

[12] Ruggeri, N., Contisciani, M., Battiston, F., and De Bacco, C. Community detection in large hypergraphs. Science Advances, 9(28), eadg9159. 2023.

[13] Schulte, S., Morillas, S., Gregori, V, and EE Kerre, EE. A new fuzzy color correlated impulse noise reduction method IEEE Transactions on Image Processing 16 (10), 2565-2575, 92. 2007.

[14] Shao, J., Han, Z., Yang, Q., and Zhou, T. Community detection based on distance dynamics. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1075-1084). 2015.

[15] Su, X., Xue, S., Liu, F., Wu, J., Yang, J., Zhou, C., ... and Philip, S. Y. A comprehensive survey on community detection with deep learning. IEEE Transactions on Neural Networks and Learning Systems. 2022.

[16] Wu, W., Zhang, H., Zhang, S., and Witlox, F. Community detection in airline networks: An empirical analysis of american vs. southwest airlines. Journal of Advanced Transportation, 2019, 1-11. 2019.

[17] You,T., Cheng,H.-M., Ning,Y.-Z., Shia, B.-C., and Zhang, Z.-Y. Community detection in complex networks using density-based clustering algorithm and manifold learning, Physica A: Statistical Mechanics and its Applications, vol. 464, pp. 221-230. 2016.