



HAL
open science

Finding Influential Nodes in Networks with Community Structure

Zakariya Ghalmane, Stephany Rajeh, Chantal Cherifi, Hocine Cherifi,
Mohammed El Hassouni

► **To cite this version:**

Zakariya Ghalmane, Stephany Rajeh, Chantal Cherifi, Hocine Cherifi, Mohammed El Hassouni. Finding Influential Nodes in Networks with Community Structure. Network Modeling, Learning and Analysis (NMLA), WorldCIST2020 Workshop, Apr 2020, Budva (Online), Montenegro. pp.3. hal-03059116

HAL Id: hal-03059116

<https://hal.science/hal-03059116>

Submitted on 13 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Finding Influential Nodes in Networks with Community Structure

Zakariya Ghalmane^{1,2}, Stephany Rajeh², Chantal Cherifi³, Hocine Cherifi², and Mohammed El hassouni¹

¹ LRIT, Mohammed V University, Rabat, Morocco,

² DISP Lab, University of Lyon 2, Lyon, France

³ LIB, University of Burgundy, Dijon, France

1 Introduction

Identifying influential nodes is a fundamental issue in complex networks. Several centrality measures take advantage of various network topological properties to target the top spreaders. However, the vast majority of works ignore its community structure while it is one of the main properties of many real-world networks. In our previous work⁴, we show that the centrality of a node in a network with non-overlapping communities depends on two features: Its local influence on the nodes belonging to its community, and its global influence on nodes belonging to the other communities. For this end, we introduced a framework to adapt all the classical centrality measures proposed for networks with no community structure to non-overlapping modular networks. We proposed a two-dimensional vector (the so-called "Modular centrality"), where each dimension accounts for a different type of influence that the nodes can exert in the network. Its first component is measured by computing the classical centrality on the Local network. This network is formed only from the intra-community links of the original network. Additionally, its second component is quantified by computing the classical centrality on the Global network. This network is formed only from the inter-community links of the original network. Depending of the strength of the community structure, these two components are more or less influential. In a recent study⁵, we extended this framework to networks with overlapping modules. Indeed, it is a frequent scenario in real-world networks where nodes usually belong to several communities, especially for social networks. The "Overlapping Modular Centrality" is a two-dimensional measure that quantifies the local and global influence of overlapping and non-overlapping nodes. The global component of this vector is defined in the same way as the Modular centrality. It is computed on the global network obtained by removing all the intra-community links from the original network. However, the local component computation depends on the nature of nodes. For a non-overlapping node, as previously, only members of its community are considered. For an overlapping node, all the communities that the node belongs to are merged in a single community. Extensive experiments have been performed on synthetic and real-world data using the Susceptible-Infected-Recovered

⁴Ghalmane, Z., El Hassouni, M., Cherifi, C., & Cherifi, H. (2019). Centrality in modular networks. *EPJ Data Science*, 8(1), 15.

⁵Ghalmane, Z., Cherifi, C., Cherifi, H., & El Hassouni, M. (2019). Centrality in complex networks with overlapping community structure. *Scientific reports*, 9(1), 1-29.

(SIR) epidemic model. Results show that the proposed framework (the Modular centrality for networks with non-overlapping modules and the Overlapping Modular Centrality for networks with overlapping modules) outperforms its standard centralities designed for non-modular networks. These investigations provide better knowledge on the influence of the various parameters setting the community structure on the nodes' centrality. Moreover, a straightforward combination of both components (modulus of the two-dimensional vector) has been evaluated. Comparative analysis with competing methods shows that they produce more efficient centrality scores.

2 Proposed approach

We restrict our attention to non-overlapping community structure. We only present the algorithm used to compute the Modular centrality. It can be summarized as follows:

- Step 1. Choose a standard centrality measure β .
- Step 2. Remove all the inter-community edges from the original network G to obtain the set of communities \mathcal{C} forming the local network G_l .
- Step 3. Compute the local measure β_L for each node in its own community.
- Step 4. Remove all the intra-community edges from the original network to reveal the set of connected components \mathcal{S} formed by the inter-community links.
- Step 5. Form the global network G_g based on the union of all the connected component.
- Step 6. Compute the global measure β_G of the nodes linking the communities based on each component of the global network.
- Step 7. Add β_L and β_G to the Modular centrality vector B_M .

3 Results

In Figure 1, we report the results of the SIR simulations using the Degree centrality. The relative difference of the outbreak size Δr between the various measures (Local component, Global component, and the modulus of either the Modular Degree centrality or the Overlapping Modular Degree centrality) and the classical Degree centrality have been computed while varying both the fraction of initial spreaders and the mixing proportion parameter μ . The relative difference of the outbreak size represents the difference between the final number of recovered nodes for the centrality measure under test and the standard centrality. It is positive if the centrality under test is more efficient than standard centrality. Moreover, the mixing parameter μ allows to control the strength of the community structure. For low values of μ , the communities are well separated (few inter-community links), while high values of μ indicates a weak community structure. Results for networks with strong community structure are illustrated in the left panel of Figure 1. We notice that in both type of networks, Δr is always positive for the local component, while it is always negative for the global component. Indeed, in networks with strong community structure the local influence predominates. The same type of results are shown in the right panel of Figure 1 for networks with weak community structure strength. This time, one can notice that Δr is always positive for the global component, while it is negative for the local component. Indeed, in this situation, the global influence takes over the local influence. Furthermore, we notice that the measure that combines both components is always the most efficient in any case.

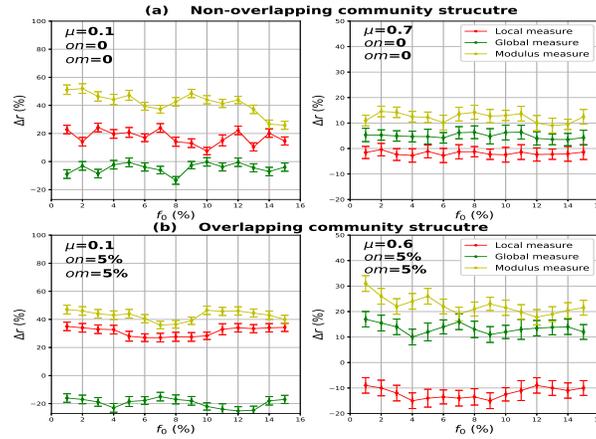


Fig. 1. Relative difference of the outbreak size Δr as a function of the portion of the initial spreaders f_0 . Where $\Delta r = (R_c - R_s)/R_s$, R_c and R_s are the final number of recovered nodes for the centrality measure under test and the standard Degree centrality. Each value is the average of 100 SIR simulations per method and fraction of initially infected nodes. Positive value of Δr means that the centrality under test is more efficient than the standard centrality. Synthetic networks generated with the LFR algorithm are used with various community structure strength μ . (a) is for networks with non-overlapping community structure. (b) is for networks with overlapping community structure, where the number of overlapping nodes on is fixed to 5% of the size of the network, and the community membership parameter om is equal to 5% of the total number of communities.

To summarize, in order to evaluate efficiently the centrality of nodes in networks with community structure (overlapping communities and non-overlapping communities), we cannot rely on the traditional centrality measures that do not take into account this topological property. It is necessary to consider and combine properly the two types of influence (local, global) occurring in such networks. In this study, we ignored the positional hierarchy of nodes in the original network. In our future work, we intend to take into account this property.