UNIVERSITATEA BABEȘ-BOLYAI
BABEȘ-BOLYAI TUDOMÁNYEGYETEM
BABEȘ-BOLYAI UNIVERSITÄT
TRADITIO ET EXCELLENTIA

Optimal community detection in multiplex and multilayer networks based on Voronoi diagrams

Szabolcs VIDÁM
Botond MOLNÁR, PhD

Report about ongoing research based on publication Community detection in directed weighted networks using Voronoi partitioning by Botond Molnár, Ildikó-Beáta Márton , Szabolcs Horvát & Mária Ercsey-Ravasz

Introduction. The surrounding world is made up from simple systems to extremely complex ones, but in the last couple of years scientists started representing and modelling these systems using complex networks. This kind of representation and studying of complex systems became very widespread in fields like biology, neuroscience, social sciences, information technology, etc. In this context an important provocation is the possibility of detecting optimal communities (clusters) in these types of complex networks. A community (cluster) in a network can be defined as the collection of nodes which have more links and/or links with higher weight between them. It is easy to see that this definition of communities is rather a qualitative one, however, up until now there is no proper mathematical definition which could be applied in this case. It is easy to see that the clustering problems becomes even more complex and difficult to solve as we move from the mathematical network models to modelling real-life systems in which – naturally – two important characteristics of the links play a crucial role, namely the direction and weight of the links. Although, these types of networks describe more accurately the complex systems they represent, generally we still tend to use the undirected and unweighted models to represent complex systems. Defining optimal communities is an extremely hard problem and falls into the category of the so-called NP-hard problems. Solving the problems becomes very inefficient on large networks.

One particularly compelling application of network clustering is in studying migration patterns. The Austrian internal migrations (2002-2022) network, for instance, provides valuable insights into how people relocate within the country over time. By clustering regions based on migration flows, researchers can identify stable communities, detect shifts in migration trends, and analyze the effects of economic or policy changes on internal movement patterns.

Beyond migration studies, clustering is widely used in other fields. In neuroscience, it helps reveal functional connectivity in brain networks, aiding in the understanding of neurological disorders. In social sciences, clustering allows for the detection of hidden communities within social and communication networks, improving for example targeted information dissemination. Similarly, in epidemiology, clustering can track the spread of diseases through human contact networks, supporting better containment strategies.

By constructing Voronoi cells around generator nodes and assigning other nodes to their closest clusters, this method offers a natural way to define community structures. Applying Voronoi-based clustering to multilayer networks allows for an analysis of how communities evolve across layers and how their structure is influenced by inter-layer dependencies. This extension is particularly relevant for real-world datasets, such as internal migration networks, where each layer corresponds to a different time snapshot, enabling researchers to study community stability and transitions over time. As the study of multilayer networks advances, developing robust, scalable, and interpretable community detection methods remains a critical challenge. The integration of Voronoi-based clustering into multilayer network analysis represents a step toward capturing the complexity of real-world systems while maintaining computational efficiency and interpretability. Recent advances have extended community detection to multilayer networks, where different layers can represent snapshots of a dynamic system or different types of interactions within the same system. Multilayer networks provide a richer and more flexible framework for studying complex systems compared to single-layer networks. By considering multiple layers simultaneously, researchers can capture temporal dynamics, structural variations, and cross-layer interactions that are otherwise lost in traditional network models.

Multilayer randomly generated benchmark networks. After showing the capabilities of the Voronoi clustering algorithm on simple weighted and directed networks we generalized and applied it to multilayer networks. First, we used the LFR algorithm to generate different types of multilayer networks: 1) same topology between layers, but different weights, no inter-layer connections; 2) different topology and weights between layers, no inter-layer links; 3) different topology and weights between layers and with inter-layer weighted links. Similarly, to the single-layer case we have randomly generated 36 000 LFR benchmark networks with $N = 1000$ nodes, average degree $\bar{k} = 100$ and the mixing parameter $\mu = 0.1, 0.2, \ldots, 0.9$, using both normal and power law distributions for the link weights. The inter- and intra-community weight distributions are varied from more distinct to more similar. In Figure 1 we show the modularity of the detected communities as a function of the weight ratio between inter- and intra-community links. Curves of different colors correspond to LFR benchmark networks generated with different mixing parameters Each data point is averaged over 100 LFR networks. We observe that while all community detection methods perform well on easy benchmark instances, they differ in how their performance decreases as the problem gets more difficult. Figure 2 shows the performance of the algorithm by presenting the mutual information between the obtained and the ground truth clustering of the benchmark networks.

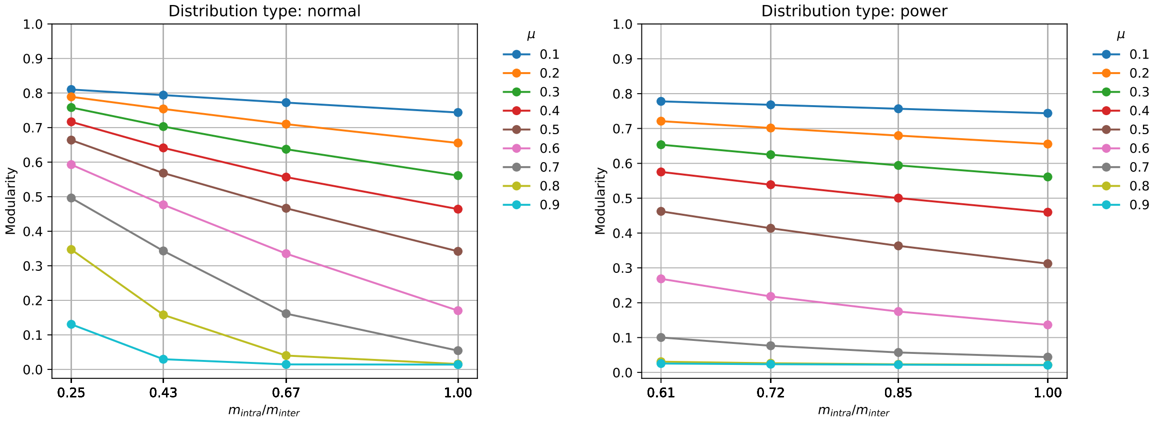

Figure 1. Performance of the Voronoi algorithm using a total of 36 000 benchmark networks. The modularity of the detected communities is shown as a function of the weight ratio between inter- and intra-community links. Curves of different colors correspond to LFR benchmark networks generated with different mixing parameters ($\mu = 0.1, 0.3, \ldots, 0.9$). All benchmark networks have $N = 1000$ nodes and mean degree $\bar{k} = 100$. The parameters of the link weight distributions were chosen for both, normal and power distributions, as follows: $(m_{inter}, m_{intra}) = (0.20{,}0.80), (0.22{,}0.78), \ldots, (0.50{,}0.50)$. Each data point is averaged over 100 LFR network.

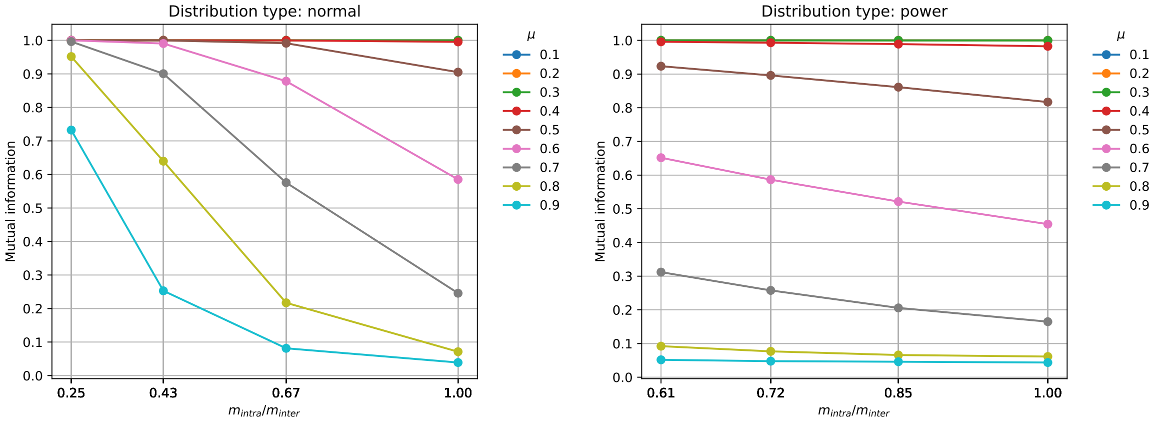

Figure 2. Performance of the Voronoi algorithm using a total of 36 000 benchmark networks. The mutual information between the detected and ground truth communities is shown as a function of the weight ratio between inter- and intra-community links. Curves of different colors correspond to LFR benchmark networks generated with different mixing parameters ($\mu = 0.1, 0.3, \ldots, 0.9$). All benchmark networks have $N = 1000$ nodes and mean degree $\bar{k} = 100$. The parameters of the link weight distributions were chosen for both, normal and power distributions, as follows: $(m_{inter}, m_{intra}) = (0.20{,}0.80), (0.22{,}0.78), \ldots, (0.50{,}0.50)$. Each data point is averaged over 100 LFR network.

It is interesting to show how the cluster structure forms in individual benchmark networks between layers. This can be shown by calculating the coherence matrix between the obtained community structure between different layers of the same network. In Figure 3 we show a benchmark network with $N = 1000$, average degree $\bar{k} = 100$, mixing parameter $\mu = 0.3$ and 5 layers using power law weight distribution ($m_{inter} = 0.3, m_{intra} = 0.7$). To illustrate even further, one can plot the network using a force layout algorithm and color the nodes according to the probability to fall in one of the existing clusters between the layers. This is shown in Figure 4 in case of the same networks we presented in Figures 16.

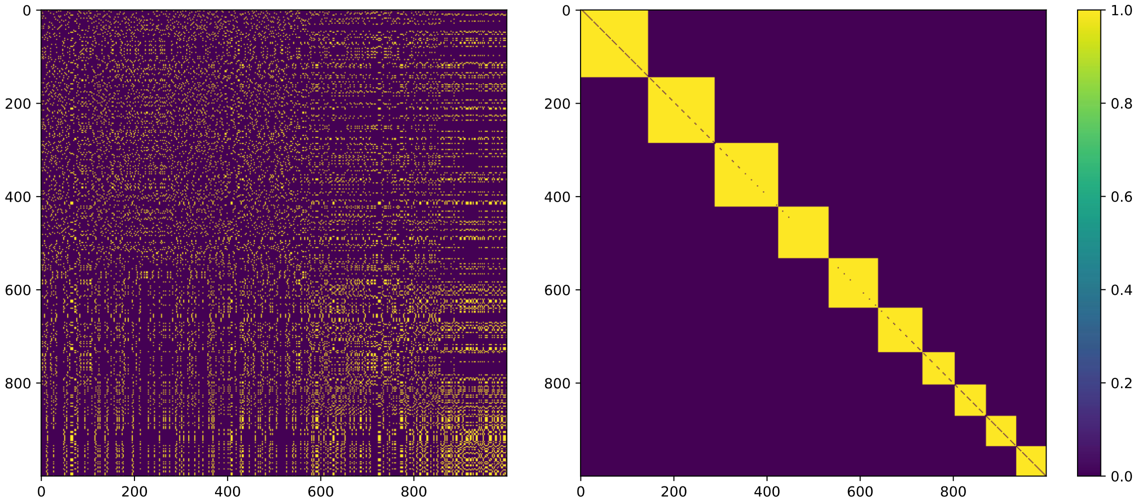

Figure 3. Coherence matrix of a clustering obtained by the Voronoi algorithm in case of a multilayer network with $N = 1000$ nodes and average degree $\bar{k} = 100$, mixing parameter $\mu = 0.3$ and 5 layers. The weights have a power law distribution with $(m_{inter}, m_{intra}) = (0.30, 0.70)$. The colors represent the probability of two nodes being partitioned in the same cluster throughout the layers, where yellow means that in every layer those nodes are partitioned in the same cluster, while dark purple corresponds to the state when two nodes are never assigned to the same cluster (see legend). On the right side we have rearranged the nodes by the coherence value starting with the largest in the top left corner.

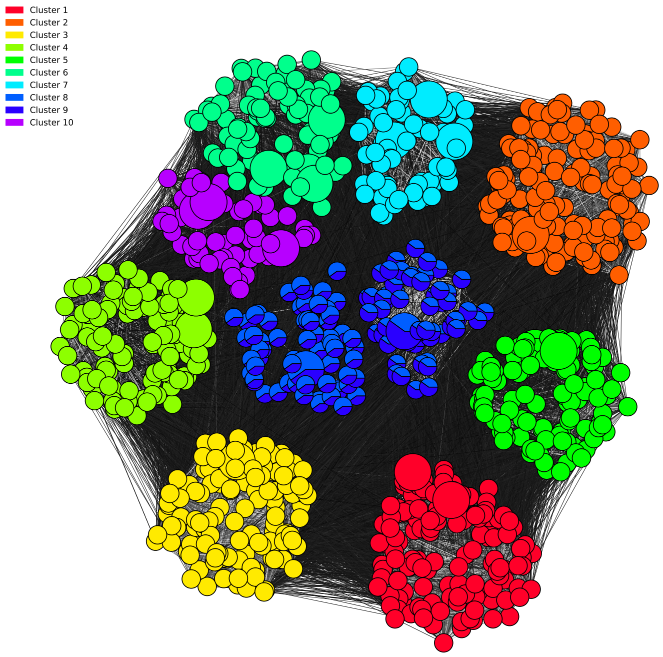

Figure 4. Force layout plot of the clustering obtained by the Voronoi algorithm in case of a multilayer network with $N = 1000$ nodes and average degree $\bar{k} = 100$, mixing parameter $\mu = 0.3$ and 5 layers. The weights have a power law distribution with $(m_{inter}, m_{intra}) = (0.30{,}0.70)$. Different colors of the nodes represent the obtained clusters. The pie chart of the node colors shows the probability of the node being assigned to either of the obtained communities.

Real-world multilayer networks. The previously studied internal migration network in case of Austria contains data for each year between 2002 and 2022. In the previous study we merged the different migration statistic data relevant for each year and performed the clustering altogether for the 20-year period, however, it would be a more accurate study to construct a multilayer network, where each layer corresponds to one year of internal migration statistics data within Austria. In this manner we end up with a multilayer network which describes the time evolution of the internal migration in Austria. In Figure 5 we show the coherence matrix for this network as well as the distance matrix between any to settlements within Austria.

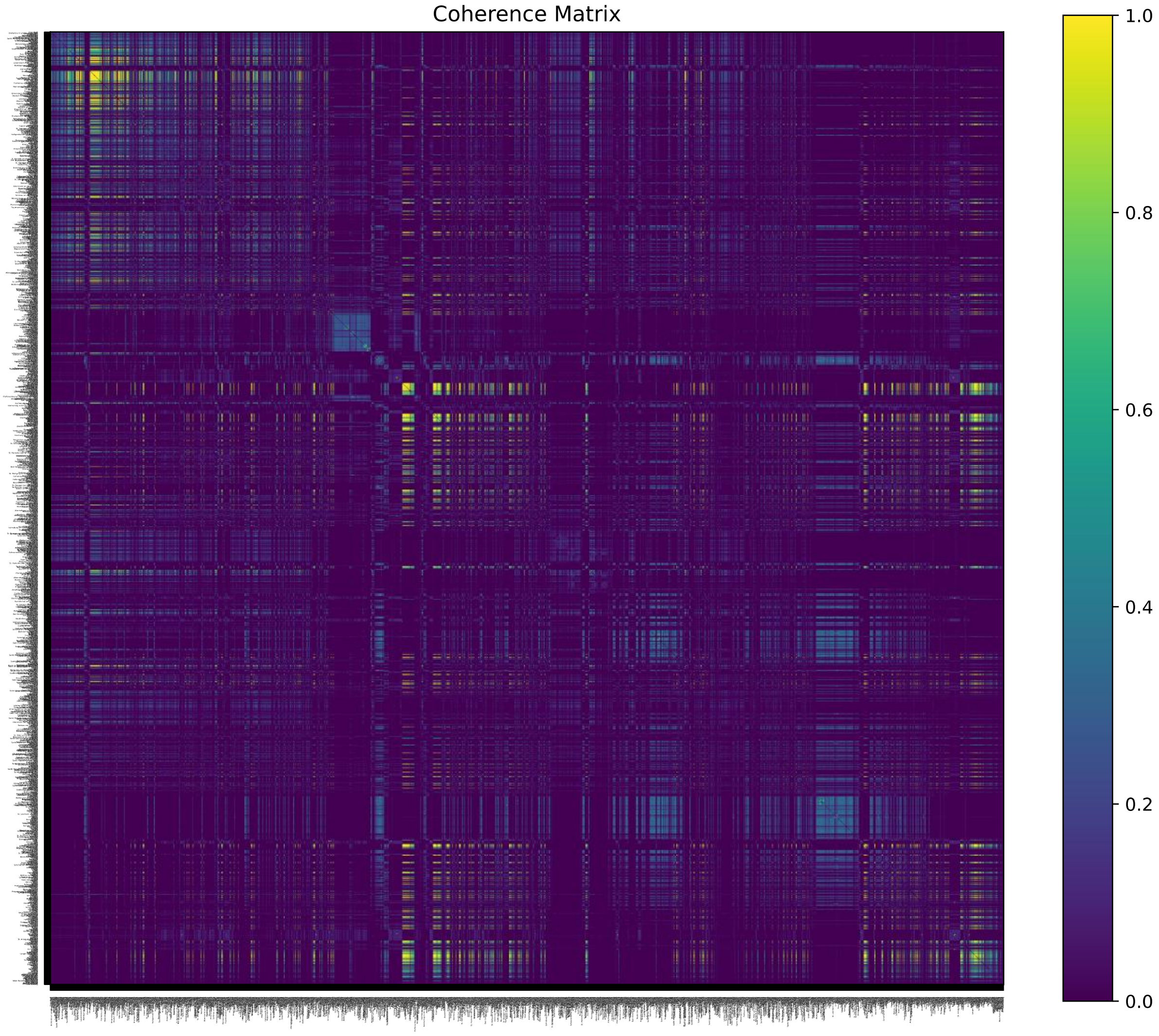

Figure 5. Coherence matrix of the clustering obtained by the Voronoi algorithm in case of the Austrian internal migration network, where each layer corresponds to one year of migration data between 2002 and 2022. The colors represent the probability of two nodes being partitioned in the same cluster throughout the layers, where yellow means that in every layer those nodes are partitioned in the same cluster, while dark purple corresponds to the state when two nodes are never assigned to the same cluster (see legend).

If we plot also the Euclidian distance between every settlement pair within Austria (Fig. 19), one can see that there is a fairly good correlation between the distances and the internal migration flaw.

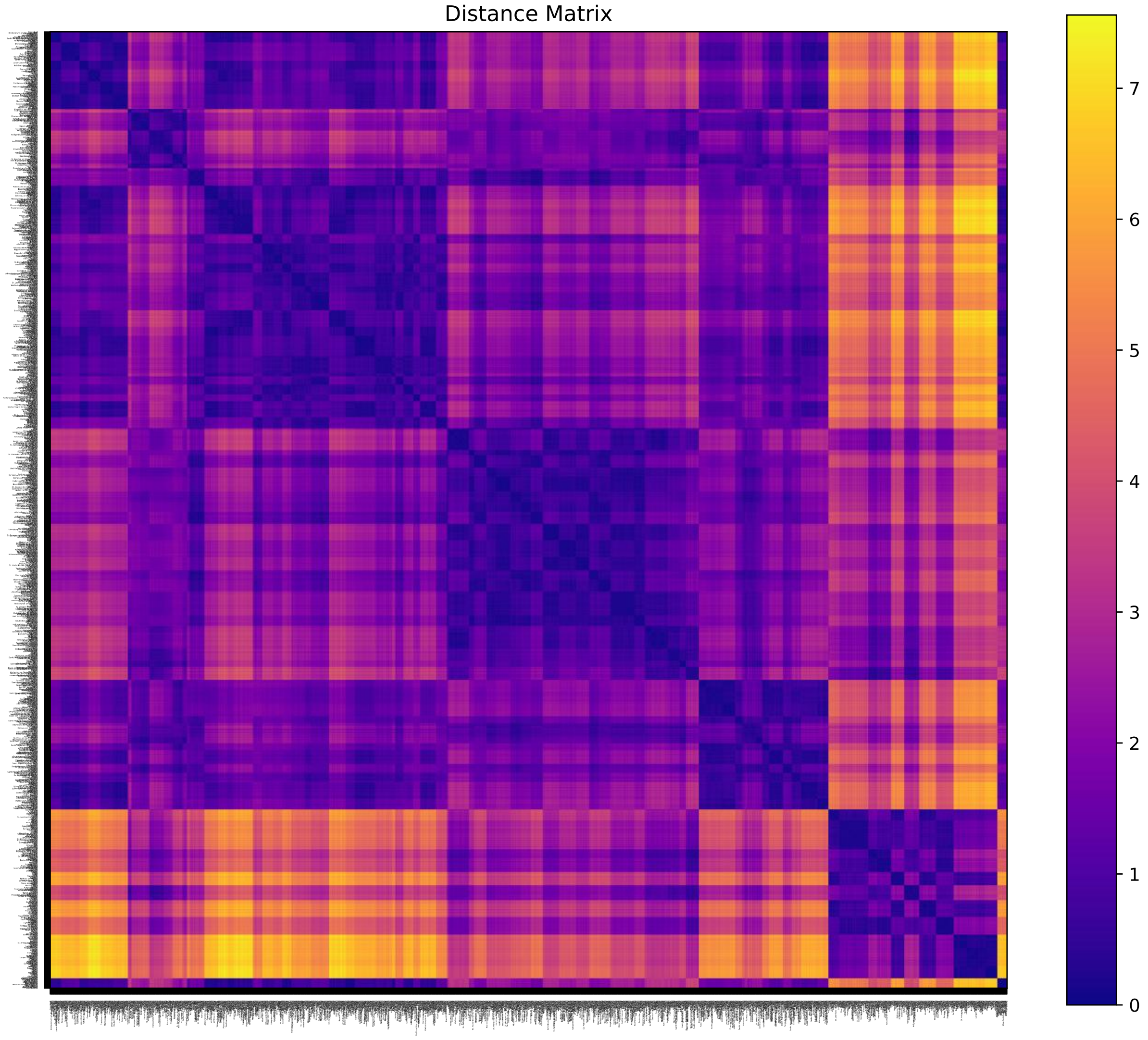

Figure 6. Euclidian distance matrix of the settlements in Austria.