Multiresolution Clustering on Massive Attributed Graphs by Means of Optimal Aggregated Markov Chains

Huijia Li, Fanghao Lou, Qiqi Wang, Guijun Li, and Matjaž Perc

***Abstract*—The efficient clustering of attributed graphs is a critical and challenging problem that has attracted much attentions across various research fields. Despite recent advancements, many challenges still exist, particularly in achieving a balance between clustering quality and computational efficiency for massive attributed graphs. To address these problems, we propose a new graph clustering algorithm based on Optimal Aggregated Markov Chains (OAMC), designed to identify an optimal partition in which the aggregated Markov chain closely approximates the behavior of a network-induced random walker. We formalize the concept of lifted aggregated transition probability matrix and utilizes the Kullback-Leibler divergence to determine the optimal partition. Given that the initial and aggregated transition probability matrices reside on different state spaces, OAMC constructs a relaxed alternative space that contains partition information and employs a multiplicative optimization method to minimize the KL divergence. The proposed algorithm effectively uncovers hierarchical cluster structures at multiple resolutions, governed by a single resolution parameter. The aggregated transition probability matrix provides a natural representation of block coupling strengths, facilitating the tracing of relationships between clusters across hierarchical layers. Moreover, the number of clusters can be specified based on the resolution parameter, providing an meaningful interpretable metric of the significance of cluster configuration. We validate the performance of our framework through comprehensive experiments on synthetic and real-world benchmark graphs. The results demonstrate that OAMC outperforms state-of-the-art graph clustering methods in terms of both efficiency and effectiveness.**

***Index Terms*—Graph clustering, attributed networks, aggregated Markov chain, KL divergence, multipulative optimization.**

Received 14 May 2025; revised 5 June 2025; accepted 10 June 2025. Date of publication 13 June 2025; date of current version 21 November 2025. The work of Matjaž Perc was supported by Slovenian Research and Innovation Agency (Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije) under Grant P1-0403. Recommended for acceptance by Prof. Chao Gao. *(Corresponding author: Matjaž Perc.)*

Huijia Li, Fanghao Lou, and Qiqi Wang are with the School of Statistics and Data Science, AAIS, LPMC and KLMDASR, Nankai University, Tianjin 300074, China (e-mail: hjli@nankai.edu.cn; 2113295@mail.nankai.edu.cn; qiqi.wang@nankai.edu.cn).

Guijun Li is with the School of Mathematical Sciences, Hebei Normal University, Shijiazhuang 050010, China (e-mail: ligj@hebtu.edu.cn).

Matjaž Perc is with the Faculty of Natural Sciences and Mathematics, University of Maribor, 2000 Maribor, Slovenia, also with Community Healthcare Center Dr. Adolf Drolc Maribor, 2000 Maribor, Slovenia, also with the Department of Physics, Kyung Hee University, Seoul 02447, Republic of Korea, also with the Complexity Science Hub, 1030 Vienna, Austria, and also with University College, Korea University, Seoul 02841, Republic of Korea (e-mail: matjaz.perc@gmail.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSE.2025.3579427, provided by the authors.

Digital Object Identifier 10.1109/TNSE.2025.3579427

I. Introduction

MANY real-world network systems contain numerous highly interconnected entities, and each is characterized by diverse attribute information [1], [2]. These systems are often effectively modeled as attributed graphs, where nodes represent users, attributes capture user profiles, and edges denote interactions. A fundamental and challenging task is to partition such attributed graphs into meaningful clusters that exhibit cohesive topological structures along with similar node attributes, known as *Attributed Graph Clustering (AGC)* [3], [4]. AGC has a broad range of applications across multiple domains, including the Internet, protein-protein interaction networks, trade networks, and scientific collaborations [5].

Despite notable progress in graph clustering research, several key challenges remain open and unresolved. First, the integration of topological and attribute information is inherently complex, as these two aspects often represent heterogeneous data sources. Existing approaches, such as linear combination techniques and model-based methods [6], [7], face difficulties in parameter tuning, defining meaningful distance functions, and selecting appropriate priors, which often necessitate rigorous interpretation of attribute clusters. Second, many graph clustering algorithms predominantly focus on topological information while overlooking latent hierarchical structures [8]. A thorough understanding of the multiresolution organization within attributed networks is crucial, particularly given the intricate membership dynamics in social settings. Lastly, AGC algorithms are often computationally expensive, especially when multiple attributes are simultaneously considered. Achieving an optimal balance between clustering effectiveness and computational efficiency remains a challenging and open research problem.

To address these problems, in this paper, we present a new insight of attribute graph clustering via random walk theory and propose an efficient graph clustering algorithm based on **O**ptimal **A**ggregated **M**arkov **C**hains (OAMC), leveraging multiplicative optimization of the KL divergence on augmented attributed networks. The flowchart of OAMC is shown in Fig. 1 and the primary contributions are three-folds as follows:

1) We introduce an innovative approach for identifying an optimal partition such that the aggregated Markov chain replicates the behavior of the network-induced random walker as closely as possible. This is accomplished by formalizing the concept of μ-lifting Markov chain on

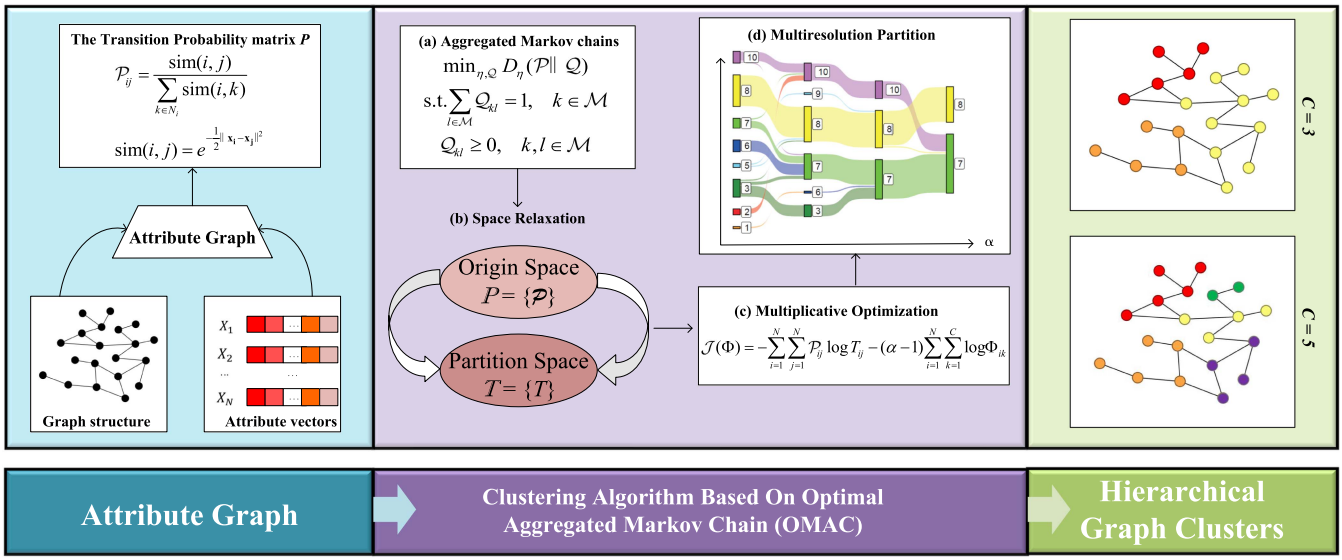

Fig. 1. The flowchart of the attributed graph clustering based on Optimal Aggregated Markov Chain (OAMC). First, we extend the attributed graph data to a Markov augmented attributed network with a transition probability matrix $\mathcal{P}$. The OAMC algorithm proceeds through four main steps: (1) proposing the KL divergence between original and aggregated transition probability matrix as the objective function, (2) introducing a new relaxed solution space for partition, (3) performing multiplicative optimization to effectively minimize the KL divergence, and (4) tuning the resolution parameter α to find the multiresolution partition. Finally, the multiresolution cluster configurations are obtained.

attribute augmented graph and minimizing the KL divergence between the original and aggregated transition probability matrices $\mathcal{P}$ and Ω. The aggregated transition probability matrix Ω naturally represents the coupling strength between clusters.

2) Minimizing the KL divergence $D_\eta(\mathcal{P}\|\Omega)$ directly is very difficult due to the different state spaces of the original transition probability matrix $\mathcal{P}$ and the aggregated transition probability matrix Ω. To address this problem, OAMC constructs a relaxed solution space for Ω that incorporates partition information, facilitating multiplicative optimization to effectively minimize the KL divergence. Moreover, the number of clusters can be specified by adjusting the resolution parameter α, which serves as a measure of the significance of a given cluster configuration.
3) The proposed algorithm reveals hierarchical cluster structures across multiple resolutions, controlled by parameter α. By adjusting α, OAMC can decompose the network into finer or coarser clusters, with larger α values leading to more granular partitions. OAMC directly utilizes the aggregated transition probability matrix Ω to trace coupling relationships between clusters across hierarchical layers, providing a clear interpretation of the network's multi-level organization.

The proposed algorithm is highly efficient: the time complexity is among the fast methods for sparse networks. Extensive simulations using both synthetic benchmark networks and real-world networks are finally conducted to verify the algorithmic performance. The remainder of this article is organized as follows: Section II reviews relevant literature and Section III presents the problem formulation of the framework. Section IV provides the details of the multiplicative optimization algorithm, and Section V analyzes the key properties of OAMC. Experimental results on multiple datasets are presented in Section VI. Finally, Section VII concludes the framework and provides directions for future research.

II. Previous Related Literature

A. Traditional Graph Clustering Methods

Over the past few decades, a number of optimization techniques for *Modularity* have been developed, which measures the quality of cluster configurations. Among the most effective methods are Louvain [9] and GLEAM [10], both capable of processing large-scale networks with billions of nodes. However, Modularity optimization suffers from a severe limitation referred as the "resolution limit" [11], which hinders the identification of fine clusters. To address this, FAMCS [12] integrates center locating and membership optimization to efficiently find partitions in plain graphs, overcoming the resolution limit. Additionally, SCD [13] partitions graphs by maximizing weighted community clustering, making it applicable to various types of weighted networks. To further overcome the effects of nonlinear properties on graph clustering, sE-Autoencoder [14] combines the nonlinear representation of autoencoders and evolutionary clustering, which is a pioneering work in graph clustering. And to enhance clustering performance in challenging multiplex graphs with sparse connectivity and noisy structures, Li et al. [15] propose an interesting and promising method named CSGR by incorporating consensus prior information into a graph regularization strategy.

B. Markov Based Clustering

Some graph clustering methods adopt multiagent system techniques, such as Markov process [16], based on the intuition that

agents tend to remain within a cluster due to higher internal link density [17]. Notable among these is Infomap [18], which leverages random walks to capture system-wide dynamics from local agent interactions, using refined Modularity optimization to detect clusters. Another Markov based algorithm is Walktrap [19], known for its effectiveness in hierarchical clustering for attributed networks. Spectral graph cut models with random walk, like Metis [20], solve the minimum cut problem using a straightforward C-way partitioning heuristic, while Cluto [21] employs an agglomerative hierarchical clustering method with a two-phase heuristic strategy. BigClam [22], a prominent overlapping clustering method, assigns each node a membership vector and combines random walk with likelihood maximization to efficiently identify overlapping clusters. E. Akbas et al. proposed an attribute-aware random walks based method AA-Cluster [23], which employs truncated random walks to learn the latent representations for attributed graphs. However, the above methods usually suffer from a dynamic process for increasing the inhomogeneity, which will directly affect the accuracy and computational cost. To address this problem, Wang et al. [24] propose a pioneering benchmark for evolutionary Markov based clustering, which optimizes the dynamic process between stochastic flows and transition probabilities through a Physarum-inspired feedback relationship, which has advanced the reserach scenarios of Markov based clustering.

C. Attributed Graph Clustering

Different with traditional graph clustering methods which focus solely on connectivity, recent methods have sought to address this by integrating topological and attribute data. Optimization is the most common way for attributed graph clustering, for example, CESNA [25] adopts a statistical significant approach to uncover attributed clusters, while EDCAR [26] uses subspace latent optimization technology. SA-Cluster [27] applies the distance-based clustering and augments the graph to enhance structural and attribute similarities. Nevertheless, challenges such as defining effective distance functions and selecting suitable priors still exist. To overcome these limitations, Li et al. [28] proposes MOEA-SA, a multi-objective evolutionary algorithm that optimizes both structural and attribute similarities effectively. Recently, some deep learning algorithms are also proposed to serve attributed graph clustering due to the superior representation learning effect. For example, Zheng et al. proposed an attributed graph clustering method named CLAGC [29] under the contrastive mechanism, and multilevel cluster-preserving augmentation with self-supervised metrics are jointly optimized. Yang et al. proposed a fuzzy algorithm [3] which encodes network structures and node attributes into a compact representation with graph convolution. The framework proposed by Li et al. [30] is a very insightful and remarkable work in the field of multi-view and multimodal research. It introduces an effective approach for learning latent relationships among heterogeneous modalities through innovative manifold learning techniques and can boost the weak features and achieved the state-of-the-art performance which is a groundbreaking work in multiview clustering.

TABLE I
THEY SYMBOLS USED IN THE PAPER

symbol	description
$\mathcal{G}$	The attribute augmented graph
C	The number of clusters in $\mathcal{G}$
π	The stationary distribution
$\mathbb{P}$	The original solution space
$\mathbb{T}$	The alternative solution space
$\mathcal{P}$	The original transition probability matrix
Ω	The aggregated transition probability matrix
η	The partition function
N_i	The neighborhood set of node i
$\Omega(\eta)$	The aggregated probability ability matrix with η
$\zeta(j)$	The set of states grouped with state j
Φ_{ik}	The probability of node i belonging to cluster k
$p_t(i)$	The likelihood of sitting on i at time t
$\mathbf{x_i}$	The attribute vector of node i
$\mathrm{sim}(i,j)$	The Euclidean distance among two vectors
$D_\eta(\mathcal{P}\|\Omega)$	The KL divergence between $\mathcal{P}$ and Ω with η
$\mathcal{J}$	The cost function of KL divergence
α	The resolution parameter

Algorithm 1: Calculate the Transition Probability $\mathcal{P}_{ij}$.

Input: The attributed augment Graph $\mathcal{G} = (V, E)$, the node attribute vectors $\Lambda = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N}\}$
Output: the transition probability $\mathcal{P}_{ij}$ for each edge
1: For $i = 1 : N$
2: For $n = 1 : N$
3: Calculate attribute similarity $\mathrm{sim}(i, j)$ for each edge as $\mathrm{sim}(i, j) = e^{-\frac{1}{2}\|\mathbf{x_i}-\mathbf{x_j}\|^2}$,
4: Calculate transition probability $\mathcal{P}_{ij}$ using attribute similarity as $\mathcal{P}_{ij} = \frac{\mathrm{sim}(i,j)}{\sum_{k\in N_i}\mathrm{sim}(i,k)}$,
5: end
6: end

III. PROBLEM FORMULATION

A. Attribute Augmented Graph

We represent an attributed network as $\mathcal{A} = (V, E, \Lambda)$, where V is the set of N nodes, E is the set of L edges, and $\Lambda = \{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$ is the collection of attribute vectors associated with each node, characterizing their properties. Each node $v_i \in V$ is associated with an η-dimensional attribute vector $\mathbf{x_i} = [x_{i1}, \ldots, x_{i\eta}]$, where $|V| = N$. The symbols and variables are listed in Table I. For the attributed graph clustering, there is the following definition.

Definition 1. (Attributed Graph Clustering): The goal of attributed graph clustering is to partition the network $\mathcal{G}$ into r subsets $\mathcal{G}_i = (V_i, E_i, \Lambda)$ such that $V = \bigcup_{i=1}^{r} V_i$ for any i. The ideal attributed graph clustering result should satisfy the following two fundamental conditions: (1) nodes within the same cluster are densely connected, while nodes across different clusters have sparse connections; (2) nodes within the same cluster have similar attribute vectors, whereas nodes in different clusters have different attribute vectors.

In attributed graphs, effective clustering requires balancing structural cohesion and attribute similarity within clusters. Random walk [16] is effective for clustering as they trap a walker

within densely connected clusters characterized by high attribute similarity [31]. However, leveraging random walk theory to uncover attribute cluster configurations is a challenging topic and lack of research. To address it, we define an *attribute augmented graph*, i.e. $\mathcal{G} = (V, W)$, that incorporates Markov properties, where the edge weight w_{ij} between adjacent nodes i and j represents their attribute similarity, quantified by the Euclidean distance between their feature vectors.

$$w_{ij} = \mathrm{sim}(i,j) = e^{-\frac{1}{2}\|\mathbf{x_i}-\mathbf{x_j}\|^2} \tag{1}$$

where $\|\mathbf{x_i} - \mathbf{x_j}\|$ is the Euclidean distance defined as $\sqrt{(\mathbf{x_i}-\mathbf{x_j})\cdot(\mathbf{x_i}-\mathbf{x_j})}$. The similarity measure $\mathrm{sim}(i,j)$ ranges from 0 to 1, increasing with the similarity between two attribute vectors. Based on Markov theory, we then define the transition probability $\mathcal{P}_{ij}$ between two adjacent nodes using this attribute similarity.

$$\mathcal{P}_{ij} = \frac{w_{ij}}{\sum_{k\in N_i} w_{ik}} = \frac{\mathrm{sim}(i,j)}{\sum_{k\in N_i} \mathrm{sim}(i,k)} \tag{2}$$

in which N_i is the neighborhood set of i in the graph. The detailed procedures are shown in Algorithm 1. Notably, the larger similarity between two attribute vectors, the larger transition probability $\mathcal{P}_{ij}$.

Based on the attributed augmented graph and the transition probability $\mathcal{P}_{ij}$ defined above, the state transition of a random walk is given by

$$p_t(i) = \sum_{j=1}^{N} \mathcal{P}_{ij} p_{t-1}(j), \tag{3}$$

where $p_t(i)$ is the likelihood that a random walker moves on i at time t. Due to the transition probability $\mathcal{P}_{ij}$, the random walker tends to move to nodes with similar attribute values in the next step.

B. Optimal Aggregated Markov Chain

Consider a random walker on attribute augmented graph $\mathcal{G}$, governed by the transition probability matrix $\mathcal{P}$, as defined in (2). This network-induced random walker follows the structure of $\mathcal{G}$. We construct an aggregated Markov chain with a transition matrix $\Omega(\eta)$, which incorporates partition function η and represents a 'coarse-grained' random walk across clusters. The proposed framework aims to identify a partition η such that the stochastic properties of the aggregated Markov chain resemble those of the network-induced random walker, thereby approximating the true underlying structure of the graph. To achieve this, we introduce the concept of μ-lifting for the aggregated transition probability matrix Ω.

Definition 2: Let Ω be an aggregated transition probability matrix defined over the 'coarse-grained' state space $\widehat{S}$. Given a partition η on the 'ordinary' state space S and a probability distribution μ on the transition matrix $\mathcal{P}$, the μ-lifting of Ω with respect to the partition η is defined as follows:

$$\widehat{\Omega}_{ij}^{(\mu)}(\eta) = \frac{\mu_j}{\sum_{k\in\zeta(j)}\mu_k}\Omega_{\eta(i)\eta(j)}, \quad i,j \in S, \tag{4}$$

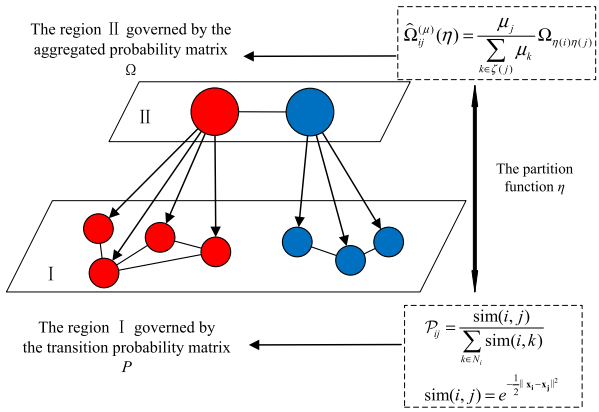

Fig. 2. The illustration of the definition of aggregated transition probability $\widehat{\Omega}_{ij}^{(\mu)}(\eta)$. Here, μ is lifting parameter, η is the partition function, $\zeta(j)$ represents the set of states grouped together with the j-th state according to the partition function η.

where $\zeta(j) = \eta^{-1}\circ\eta(j) \subset S$ represents the set of states grouped with the j-th state based on the partition function η.

The detailed definition of Ω is illustrated in Fig. 2. The primary goal of this study is to determine an optimal partition η and a corresponding aggregated transition matrix Ω that maximizes the alignment between the aggregated Markov chain and the network-induced Markov chain. This alignment is quantified by the KL divergence rate $D(\mathcal{P}\|\Omega)$, where a lower value of $D(\mathcal{P}\|\Omega)$ indicates greater similarity between $\mathcal{P}$ and Ω, suggesting that the aggregated Markov chain effectively captures the behavior of the network-induced random walker. Using the μ-lifting of Ω, the KL divergence rate is formally defined as follows:

Definition 3: Let $\mathcal{P}$ and Ω be the transition probability matrices of two Markov chains defined on state spaces S and $\widehat{S}$, respectively. The KL divergence rate is given by

$$D_\eta(\mathcal{P}\|\Omega) := \min_{\mu\in\mathcal{P}(S)} D\left(\mathcal{P}\|\widehat{\Omega}^{(\mu)}(\eta)\right), \tag{5}$$

where $\widehat{\Omega}^{(\mu)}(\eta)$ represents the μ-lifting of Ω associated with the partition function η.

In (5), the parameter μ is crucial for determining the KL divergence. Interestingly, as established in the following theorem, the KL divergence rate is minimized when μ is the stationary distribution π of $\mathcal{P}$.

Theorem 1: Let $\mathcal{P}$ and Ω be transition probability matrices corresponding to two Markov chains on state spaces S and $\widehat{S}$, respectively. For any partition function $\eta : S \mapsto \widehat{S}$, the stationary distribution π of $\mathcal{P}$ minimizes (5). The following is the minimal KL divergence rate:

$$D_\eta(\mathcal{P}\|\Omega) = D\left(\mathcal{P}\|\widehat{\Omega}^{(\pi)}(\eta)\right) = U_\eta(\mathcal{P}\|\Omega) - S_\eta(\pi), \tag{6}$$

where

$$U_\eta(\mathcal{P}\|\Omega) = \sum_{i,j\in S}\pi_i\mathcal{P}_{ij}\log\left(\frac{\mathcal{P}_{ij}}{\Omega_{\eta(i)\eta(j)}}\right),$$
$$S_\eta(\pi) = \sum_{j\in S}\pi_j\log\frac{\pi_j}{\sum_{k\in\zeta(j)}\pi_k}. \tag{7}$$

The proof of Theorem 1 is provided in Supplementary Materials. Notably, $\mu = \pi$ is not necessarily the unique minimizer of (5). The distribution μ minimizes (5) if and only if there exist constants $\{K_l, l \in \widehat{S}\}$ such that

$$\frac{\pi_j}{\mu_j} = K_l, \quad \forall j \in \eta^{-1}(l), \, l \in \widehat{S}. \tag{8}$$

Under Theorem 1, the matrix $\widehat{\Omega}^{(\mu)}(\eta)$ then coincides with $\widehat{\Omega}^{(\pi)}(\eta)$.

To determine the optimal partition $\eta : S \mapsto \widehat{S}$ and the corresponding optimal aggregated Markov chain with transition matrix Ω, we formulate the following optimization problem by minimizing $D_\eta(\mathcal{P}\|\Omega)$ as defined in (9):

$$\begin{aligned} &\min_{\eta,\Omega} \; D_\eta(\mathcal{P}\|\Omega) \\ \text{s.t.} \quad &\sum_{l\in\widehat{S}} \Omega_{kl} = 1, \quad k \in \widehat{S}, \\ &\Omega_{kl} \geq 0, \quad k, l \in \widehat{S}, \end{aligned} \tag{9}$$

where $D_\eta(\mathcal{P}\|\Omega)$ is defined in (6), and the constraints ensure that Ω is a valid Markov transition probability matrix.

With a fixed partition function η, the problem in (9) becomes convex and can be readily solved. The following theorem outlines the optimization procedure:

Theorem 2: Let η be a partition over S with corresponding state space $\widehat{S}$. For the problem in (9) with a fixed partition η, there are the following results:

(i) The optimal transition matrix $\Omega_k(\eta)$ is given by

$$\Omega_k(\eta) = \frac{\sum_{i\in\eta^{-1}(k)} \sum_{j\in\eta^{-1}(l)} \pi_i \mathcal{P}_{ij}}{\sum_{i\in\eta^{-1}(k)} \pi_i}, \quad k, l \in \widehat{S}. \tag{10}$$

(ii) The stationary distribution of the aggregated Markov chain $\Omega(\eta)$ is

$$\theta_k(\eta) = \sum_{i\in\eta^{-1}(k)} \pi_i, \quad k \in \widehat{S}. \tag{11}$$

(iii) The stationary distribution of the lifted Markov chain with transition matrix $\widehat{\Omega}^{(\pi)}(\eta)$ matches that of the original Markov chain.

The proof of Theorem 2 is presented in Supplementary Materials. While Ω can be easily computed for a given η, directly minimizing $D_\eta(\mathcal{P}\|\Omega)$ is challenging due to differences in the state spaces of $\mathcal{P}$ and Ω. The problem in (9) constitutes a mixed-integer nonlinear program, which becomes computationally expensive as the state space grows [34]. To address this question, we relax the solution space of Ω to incorporate partition information, facilitating efficient optimization using regularization techniques.

Remark 1: On a macroscopic level, the aggregated transition probability Ω_{ij} defined in (10) can be interpreted as the "coupling strength" between two clusters for a given partition η. While fine resolution transition probabilities capture localized information, coarser coupling strengths can be used to capture the global patterns by ignoring some microscopic details [1], [32].

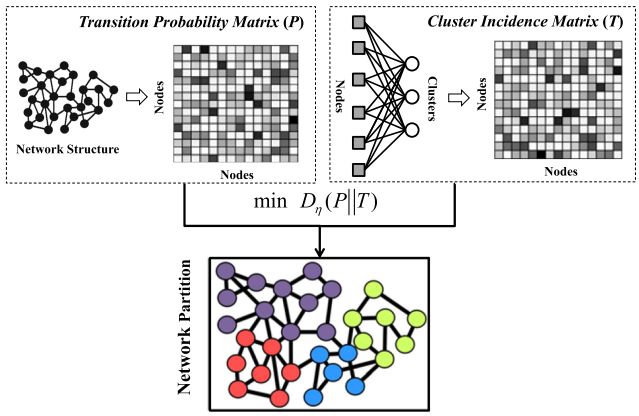

Fig. 3. The illustration of minimizing the KL divergence $D(\mathcal{P}\|T)$ between the transition probability matrix $\mathcal{P}$ and cluster incidence matrix T. Here, the metric $D(\mathcal{P}\|T)$ can be considered as a kind of *lumpability* between original and aggregated Markov chain, and "best lumped" partition is the optimal cluster configuration we aim to find out.

C. Relaxed Solution Space

As discussed before, directly minimizing $D_\eta(\mathcal{P}\|\Omega)$ to determine the optimal partition η is challenging due to the different state spaces between $\mathcal{P}$ and Ω. To resolve this problem, we introduce an alternative solution space $\mathbb{T}$ for Ω, which inherently incorporates the partition information. We then discretize this space to practically identify the optimal partition η. To define this alternative solution space, we use:

$$\begin{aligned} \mathbb{T} = \Bigg\{ T \mid T_{ij} = \sum_{k=1}^{C} \frac{\Phi_{ik}\Phi_{jk}}{\sum_{v=1}^{N} \Phi_{vk}}, \, \Phi \in \mathbb{R}^{N\times C}, \\ \sum_{k=1}^{C} \Phi_{ik} = 1, \, \forall i, \, \Phi_{ik} \geq 0 \Bigg\}. \end{aligned} \tag{12}$$

Here, $\Phi_{ik} = P(k|i)$ denotes the probability that node i belongs to cluster k, in which i, j, v represent nodes (ranging from 1 to N) and k, l represent clusters (ranging from 1 to C). Naturally, the matrix $\Phi = \{\Phi_{ik}\}$ serves as the representation of the partition function η. In a probabilistic view, T can be understood as follows: assuming a uniform prior node distribution which is represented as $P(j) = 1/N$, Bayes' theorem gives us:

$$P(j|k) = \frac{P(k|j)P(j)}{\sum_{v=1}^{N} P(k|v)P(v)} = \frac{P(k|j)}{\sum_{v=1}^{N} P(k|v)}. \tag{13}$$

Thus, the aggregated transition probability T_{ij} can be expressed as:

$$\begin{aligned} T_{ij} &= \sum_{k=1}^{C} \frac{\Phi_{ik}\Phi_{jk}}{\sum_{v=1}^{N} \Phi_{vk}} = \sum_{k=1}^{C} \frac{P(k|j)}{\sum_{v=1}^{N} P(k|v)} P(k|i), \\ &= \sum_{k=1}^{C} P(j|k)P(k|i) = P(j|i). \end{aligned} \tag{14}$$

Thus, T_{ij} can be considered as "cluster incidence matrix" which denotes the transition probability from node i to node j through a particular cluster, just as shown in Fig. 3. Notably, the matrix T

contains the partition information and has the same dimensionality as the transition matrix Ω. However, it is crucial to verify whether T and the original transition matrix $\mathcal{P}$ occupy the same solution space. Consider the solution space for the transition matrix $\mathbb{P}$, defined as:

$$\mathbb{P} = \Bigg\{ \mathcal{P} \mid \sum_{j=1}^{N} \mathcal{P}_{ij} = 1,\ \mathcal{P} = \Psi\Psi^T,\ \Psi \in \mathbb{R}^{N\times C},\ \forall i,k,\ \Psi_{ik} \geq 0 \Bigg\}. \tag{15}$$

We can deduce the following Theorem 3 and present the detailed proof on Supplementary Materials.

Theorem 3: The solution spaces $\mathbb{P}$ and $\mathbb{T}$ are equivalent, i.e., $\mathbb{P} = \mathbb{T}$.

This theorem is crucial since it establishes a connection that $\mathbb{P}$ and $\mathbb{T}$ share the same solution space. Consequently, we can replace the aggregated transition probability matrix Ω with T and redefine $D_\eta(\mathcal{P}\|\Omega)$ in (9) as the KL divergence between $\mathcal{P}$ and T, i.e., $D(\mathcal{P}\|T)$. This substitution facilitates optimization without loss of information. The formal definition of the minimization of the KL divergence between $\mathcal{P}$ and T is given by:

$$\begin{aligned} \min_{\Phi\geq 0} D(\mathcal{P}\|T) &= \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\mathcal{P}_{ij}\log\frac{\mathcal{P}_{ij}}{T_{ij}} - \mathcal{P}_{ij} + T_{ij}\right) \\ \text{subject to}\quad T_{ij} &= \sum_{k=1}^{C}\frac{\Phi_{ik}\Phi_{jk}}{\sum_v \Phi_{vk}}, \\ &\sum_{k=1}^{C}\Phi_{jk} = 1, \quad i = 1,\ldots,N. \end{aligned} \tag{16}$$

Omit the constants in $D(\mathcal{P}\|T)$, since $\sum_{i=1}^{N}\sum_{j=1}^{N} T_{ij} = N$, making the objective function equivalent to maximizing $\sum_{i=1}^{N}\sum_{j=1}^{N}\mathcal{P}_{ij}\log T_{ij}$.

Remark 2: We can define the KL divergence $D(\mathcal{P}\|T)$ as a kind of *lumpability* [33] between original and aggregated Markov chain, and if $D(\mathcal{P}\|T)$ is minimized, one can find the "best lumped" cluster configuration which contains most information of original attributed graph, just as shown in Fig. 3. In Section IV, we demonstrate that $D(\mathcal{P}\|T)$ is well-suited for optimization and present an efficient algorithm to handle the intricate structure of T. By utilizing the properties of logarithms [34], we decompose T into additive components, facilitating the optimization of the objective function via a straightforward convex-concave update rule.

IV. Optimization Algorithm

The objective function $D(\mathcal{P}\|T)$, as defined in (16), uses the matrix Φ as a parameter, where each row sums to one. Assuming the rows are generated from a common Dirichlet distribution, we control the complexity of Φ by incorporating a log-Dirichlet prior, resulting in the following cost function:

$$\mathcal{J}(\Phi) = -\sum_{i=1}^{N}\sum_{j=1}^{N}\mathcal{P}_{ij}\log T_{ij} - (\alpha-1)\sum_{i=1}^{N}\sum_{k=1}^{C}\log\Phi_{ik}. \tag{17}$$

When the elements $\mathcal{P}_{ij}$ are integers, (16) represents the log-likelihood of the following generative model: (1) sample rows of Φ from a uniform Dirichlet distribution with parameter α; (2) update each entry (i,j) based on a Multinomial distribution $\sim$ Multinomial$(\frac{1}{N}T, 1)$ for $t = 1,\ldots,t_0$. The parameter α acts as a resolution parameter for multiscale clustering: when $\alpha = 1$, the Dirichlet prior disappears, favoring larger clusters, while increasing α allows for finer granularity and more, smaller clusters.

To minimize $\mathcal{J}(\Phi)$, we employ a multiplicative update approach. Given that $\mathcal{J}(\Phi)$ is parameterized by a nonnegative matrix Φ, we first compute the gradient, which can be split into two nonnegative components, $\nabla^+_{ik} \geq 0$ and $\nabla^-_{ik} \geq 0$:

$$\nabla_{ik} \overset{\text{def}}{=} \frac{\partial\mathcal{J}}{\partial\Phi_{ik}} = \nabla^+_{ik} - \nabla^-_{ik}. \tag{18}$$

These components can typically be identified from the gradient expression, and the optimization iteratively updates Φ using the multiplicative rule: $\Phi_{ik} \leftarrow \Phi_{ik}\frac{\nabla^-_{ik}}{\nabla^+_{ik}}$ until convergence. This algorithm maintains the positivity of Φ, eliminates the need to tune the learning rate, and guarantees that $\mathcal{J}$ decreases monotonically, converging to a stable solution.

Directly applying these updates, however, is challenging due to the probability constraints on rows of Φ. Simply projecting the rows onto the probability simplex often yields suboptimal results. To address this, we introduce Lagrange multipliers $\{\lambda_i\}_{i=1}^N$ to enforce the constraints:

$$\mathcal{L}(\Phi,\lambda) = \mathcal{J}(\Phi) + \sum_i \lambda_i\left(\sum_{k=1}^{C}\Phi_{ik} - 1\right). \tag{19}$$

This formulation leads to an early version of the multiplicative update rule for Φ:

$$\Phi'_{ik} = \Phi_{ik}\frac{\nabla^-_{ik} - \lambda_i}{\nabla^+_{ik}}, \tag{20}$$

with the gradient components defined as:

$$\frac{\partial\mathcal{J}}{\partial\Phi} = \underbrace{\left[(\Phi^T Z\Phi)_{kk}s_k^{-2} + \Phi_{ik}^{-1}\right]}_{\nabla^+_k} - \underbrace{\left[2(Z\Phi)_{ik}s_k^{-1} + \alpha v_{ik}^{-1}\right]}_{\nabla^-_k}, \tag{21}$$

where $Z_{ij} = \mathcal{P}_{ij}/T_{ij}$ and $s_k = \sum_{v=1}^{N}\Phi_{vk}$. To ensure $\sum_k \Phi'_{ik} = 1$ and determine λ_i, we have:

$$\lambda_i = \frac{b_i - 1}{a_i}, \tag{22}$$

where

$$a_i = \sum_{l=1}^{C}\frac{\Phi_{il}}{\nabla^+_{il}}, \quad b_i = \sum_{l=1}^{C}\Phi_{il}\frac{\nabla^-_{il}}{\nabla^+_{il}}. \tag{23}$$

Algorithm 2: The Multiplicative Optimization Algorithm of OAMC.

Input: The transition probability matrix $\mathcal{P}$, the initialization of partition matrix Φ, the number of clusters C

Output: The optimal partition matrix Φ;

Repeat

1: $T_{ij} = \sum_{k=1}^{C} \frac{\Phi_{ik}\Phi_{jk}}{\sum_{v=1}^{N}\Phi_{vk}}$;

2: $Z_{ij} = \mathcal{P}_{ij}/T_{ij}$;

3: $s_k = \sum_{v=1}^{N}\Phi_{vk}$

4: $\nabla_k^+ = (\Phi^T Z \Phi)_{kk} s_k^{-2} + \Phi_{ik}^{-1}$;

5: $\nabla_k^- = 2(Z\Phi)_{ik} s_k^{-1} + \alpha\Phi_{ik}^{-1}$;

6: $a_i = \sum_{l=1}^{C} \frac{\Phi_{il}}{\nabla_{il}^+}$, and $b_i = \sum_{l=1}^{C} \Phi_{il}\frac{\nabla_{il}^-}{\nabla_{il}^+}$;

7: $\Phi_{ik} \leftarrow \Phi_{ik}\frac{\nabla_{ik}^- a_i + 1}{\nabla_{ik}^+ a_i + b_i}$.

Until Φ converges under the given tolerance.

Substituting this value of λ_i back into (20), we obtain:

$$\Phi_{ik} \leftarrow \Phi_{ik}\frac{\nabla_{ik}^- a_i + 1 - b_i}{\nabla_{ik}^+ a_i}. \tag{24}$$

To preserve positivity, we add b_i to both the numerator and denominator, yielding the final update rule:

$$\Phi_{ik} \leftarrow \Phi_{ik}\frac{\nabla_{ik}^- a_i + 1}{\nabla_{ik}^+ a_i + b_i}. \tag{25}$$

These steps are summarized in Algorithm 2. Notably, constructing the entire matrix T is unnecessary in practice, as the ratio $Z_{ij} = \mathcal{P}_{ij}/T_{ij}$ only involves computing the non-zero entries in S.

The following theorem guarantees the monotonicity which the algorithm observes.

Theorem 4: Denote Φ^{new} as the updated matrix for every step of Algorithm 2. We have $\mathcal{L}(\Phi^{new}, \lambda) \leq \mathcal{L}(\Phi, \lambda)$ with $\lambda_i = (b_i - 1)/a_i$.

The proof of Theorem 1 is provided in Supplementary Materials. Theorem 4 proves that Algorithm 2 not only minimizes the approximation error but also guarantees that rows converge to the probability simplex. The algorithm dynamically selects the appropriate Lagrange multipliers. Here, b_i represents the cumulative effect of unconstrained multiplicative updates, while a_i balances the force of probability simplex against the gradient.

V. Properties of the Framework

A. Initialization Setting

At the input of Algorithm 2, it is crucial to initialize the membership matrix $\Phi = \{\Phi_{ik}\}$ to ensure $\sum_{k=1}^{C}\Phi_{ik} = 1$ and $\Phi_{ik} \geq 0$ for all i, k, to avoid the emerge of trivial solutions. Initially, we set:

$$\Phi_{ik} = \frac{1}{C} + y_{ik};\ i \in \{1, 2, \ldots, N\},\ k \in \{1, 2, \ldots, C\},$$
$$\text{s.t.} \quad \sum_{i=1}^{N} y_{ik} = 0, \tag{26}$$

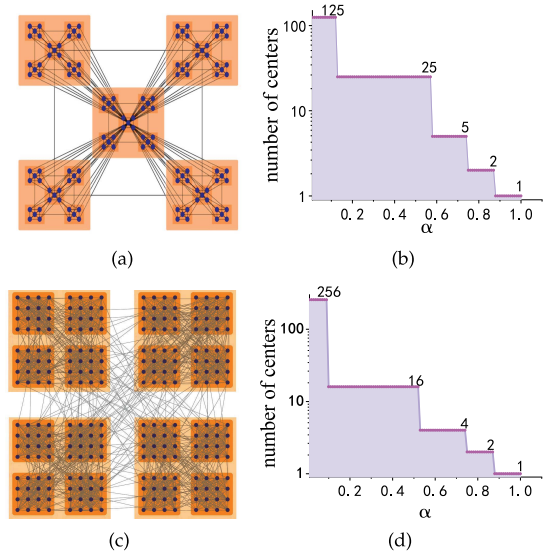

Fig. 4. The illustration of relationship between the number of clusters C with the parameter α. The subgraph (a) illustrate the RB125 network, which contains 25 small modules and 5 large modules. The change of α with number of clusters C of RB125 network is shown in subgraph (b). The subgraph (c) illustrate the H13-4 network, which contains 16 small modules and 4 large modules. The change of α with number of clusters C of H13-4 network is shown in subgraph (d).

where y_{ik} is a small Gaussian noise. The process aims to enhance the initialization performance cumulatively, which is better than a fully uniform initialization with equal values.

B. Optimal Number of Hierarchical Clusters

For hierarchical clusters, determining the optimal number of clusters, C, is fundamental question for many graph clustering algorithm. In our framework, a single resolution parameter, α, controls the hierarchical structure of the clusters. As α decreases, OAMC tends to produce more while smaller clusters. We evaluate the influence of α using two multi-level hierarchical benchmark datasets: the $RB125$ network [35] and the $H13$-4 network [36], as shown in Fig. 4(a) and (c), respectively. The persistence of cluster configurations as α varies suggests an appropriate number of clusters. For the $RB125$ and $H13$-4 networks, 25 and 16 clusters, respectively, exhibit the longest persistence, as seen in Fig. 4(b) and (d). Additionally, coarser hierarchical levels are observed with 5 and 4 clusters. As α increases, smaller clusters merge into larger modules. The persistence of parameter α related to C clusters can be considered as a kind of "significance" measure, which is used to quantify the complexity of transition from C clusters to $C-1$ clusters. A stable C over a long range of α indicates a robust cluster configuration. Therefore, we can utilize the longest range of α to determine the optimal number of clusters.

C. Multiresolution Partition and Fuzzy Clustering

The OAMC algorithm provides an optimal membership matrix $\Phi = \{\Phi_{ik}\}$ in probabilistic terms, allowing for the natural identification of overlapping clusters. In (26), the uniform initialization of $\Phi = \{\Phi_{ik}\}$ ensures that each node i is assigned

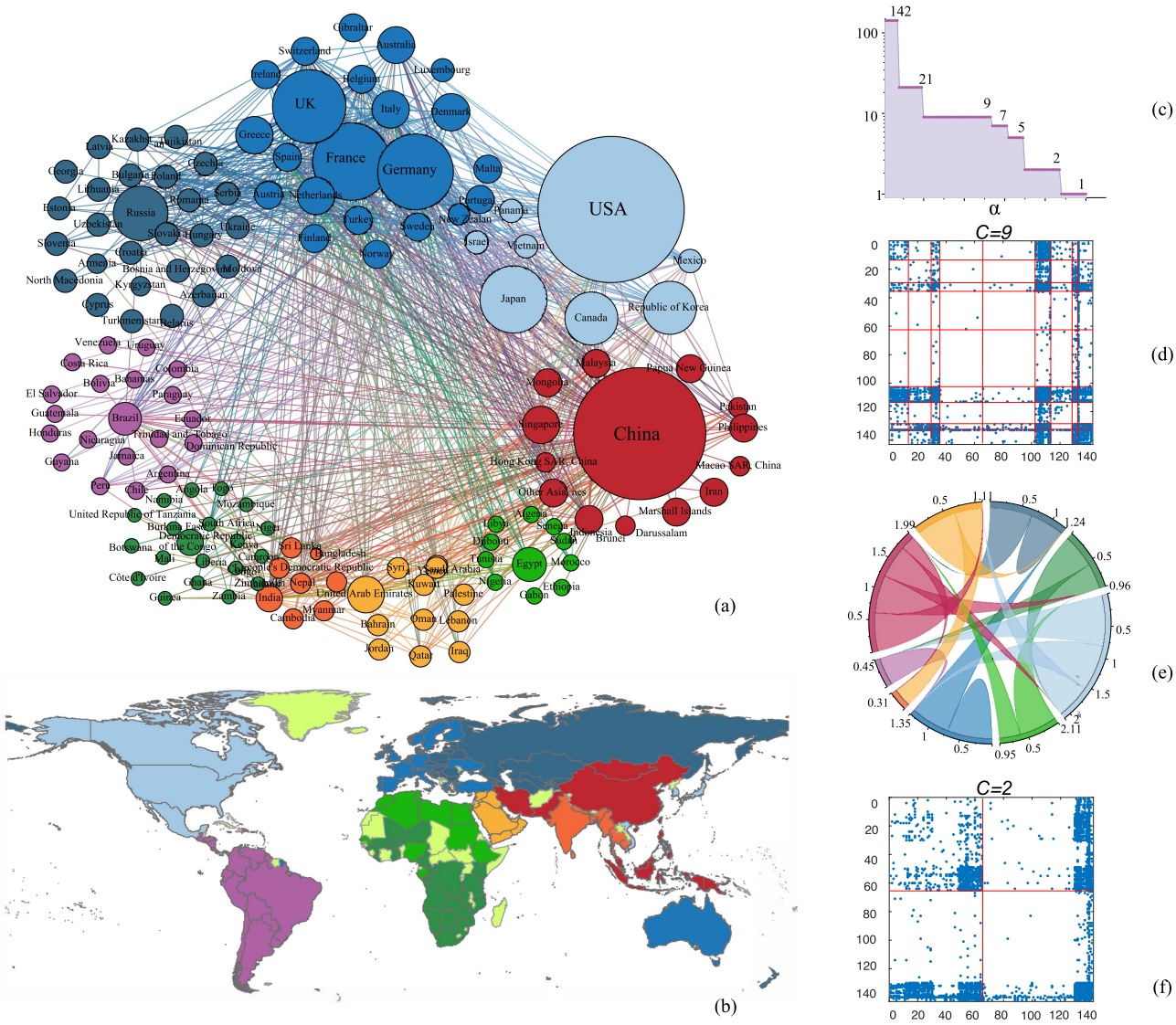

Fig. 5. Analysis of the World Trade Network using OAMC. (a) Network partition with 9 clusters. (b) Geographical representation of cluster relationships. (c) The change of number of clusters C with resolution parameter α. (d) The illustration of adjacent matrix for 9 clusters. (e) The illustration of Ω matrix for the 9-cluster division. (f) The illustration of adjacent matrix for 2 clusters.

to the cluster with the largest Φ_{ik}. If a node belongs to multiple clusters with nearly equal Φ_{ik} values, it can be considered an outlier, or a "double-dealer" in practical scenarios.

The OAMC framework also employs a single hyperparameter, α, to control the accuracy of cluster decomposition for a specific partition η. The hierarchical organization of clusters can be derived using this property. Once a partition η is obtained, the aggregated transition probability Ω_{ij}, as defined in (10), quantifies the "coupling strength" between two clusters. Fine resolution transition probabilities capture local structures, while coarser resolutions reveal broader patterns by omitting minor details [1], [32]. Furthermore, the hierarchical relationships between clusters in different layers, such as parent-child connections, can be identified by tracing nodes across layers, effectively representing these relationships as bundles of traces from child to parent clusters.

D. Time Complexity

The OAMC algorithm starts by computing the transition probability $\mathcal{P}_{ij}$ in the attributed augmented graph (Algorithm 1), followed by multiplicative optimization (Algorithm 2). In Algorithm 1, calculating $\mathcal{P}_{ij}$ involves a time complexity of $O(L)$, where L is the number of edges. The primary computational cost in Algorithm 2 comes from updating T_{ij}, ∇_k^+, ∇_k^-, and Φ_{ik}. Specifically, updating T_{ij} has a time complexity of $O(CN^2)$, where N and C represent the number of nodes and clusters, respectively. Calculating ∇_k^+ and ∇_k^- requires inverting the matrix Φ, leading to a time complexity of $O(N^2 \log N)$. Updating Φ_{ik} itself incurs a complexity of $O(N^2)$. Therefore, for sparse networks, the overall time complexity of Algorithm 2 is dominated by calculating ∇_k^+ and ∇_k^-, making the total time complexity for OAMC $O(N^2 \log N)$.

VI. Experiments

In this section, we applied OAMC to the representative World Trade Network (WTN) to evaluate its performance, particularly in analyzing hierarchical clustering structures. All experiments were conducted on a system with Dual Socket P+ (LGA-4189) 3 rd Gen Intel Scalable Processors, an Intel C621 A Chipset, and up to 8 TB Intel Optane Persistent Memory.

The WTN dataset, derived from the BACI database of CEPII[1], documents trade relationships between global regions in 2022. We computed the total trade value between each pair of regions, retaining only relationships exceeding 2 million USD. This reduced the network to 142 nodes (regions) and 1496 edges (trade relations). We used the longitude of the national capital as a node attribute to reflect geographical location.

After applying OAMC, we analyzed the resolution parameter α. Fig. 5(a) shows the network partitioned into 9 clusters, with colors representing different clusters and node size indicating trade volume. We visualized this partition on a world map in Fig. 5(b), which reveals clear regional patterns, such as NAFTA, the EU, the Middle Eastern Arab countries, and Southeast Asia, showcasing the effectiveness of OAMC. Fig. 5(c) indicates that the optimal number of clusters is 9 or 2, as evidenced by the longest duration of stability for α. To further validate the partitioning, Fig. 5(d) presents the matrix structure, highlighting significant intra-cluster connectivity and cohesiveness.

The aggregated transition probability matrix Ω, a key feature of OAMC, captures the coarse-grained coupling strength between clusters, quantifying trade interconnectivity among economic regions. Fig. 5(e) displays the Ω matrix for the 9-cluster partition, indicating stronger trade relationships among geographically and economically similar clusters (e.g., the bright blue and dark blue clusters). These relationships imply balanced or mutually beneficial economic interactions. In contrast, the red and light blue clusters exhibit numerous trade links with other regions, reflecting the dominant economic influence of the USA and China compared to their neighboring regions.

Several interesting patterns also emerged from the clustering results: (1) The red cluster maintains trade relations with both green clusters, likely due to China's strong trade ties with several African countries. (2) The two green clusters, both in Africa, exhibit cooperative trade despite differences in economic development levels. (3) The yellow cluster has trade connections with several major organizations, corresponding mainly to Middle Eastern countries in the Arabian Peninsula, which possess valuable resources like petroleum, driving their trade relations.

Further, we partitioned the WTN into 2 clusters, as shown in Fig. 5(f). This analysis revealed notable trends in capital flows, suggesting that international trade predominantly follows regional dynamics, as internal flows were slightly larger than flows between clusters. This indicates the influence of geographical proximity on trade, providing insights into potential shifts in the future global trade.

[1]http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=37

VII. Conclusion

In this paper, we introduced an efficient attributed graph clustering algorithm based on Optimal Aggregated Markov Chains (OAMC). We formalized the concept of μ-lifting for aggregated transition probability matrix Ω and employed the KL divergence between the original and aggregated transition matrices to determine the optimal partition. OAMC reveals hierarchical cluster structures at multiple resolutions, controlled by a single parameter, α. The aggregated transition probability matrix Ω captures block coupling strengths between clusters, enabling the direct tracing of cluster relationships across hierarchical layers. For the future works, this study opens several avenues for future exploration. First, we aim to optimize the current framework to handle noisy relational data where attributes may be incomplete or inaccurate. Second, we plan to extend the framework to identify attribute cluster structures on high-order network data, such as signed network, multilayer network and hypergraph. Lastly, we intend to develop an incremental clustering approach to accommodate dynamic relational data, enhancing the adaptability of OAMC in evolving network scenarios.

References

[1] C. Xiao, X. Xu, Y. Lei, K. Zhang, S. Liu, and F. Zhou, "Counterfactual graph learning for anomaly detection on attributed networks," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 10, pp. 10540–10553, Oct. 2023.

[2] M. Qin, C. Zhang, B. Bai, G. Zhang, and D. Y. Yeung, "Towards a better tradeoff between quality and efficiency of community detection: An inductive embedding method across graphs," *ACM Trans. Knowl. Discov. Data*, vol. 17, no. 9, pp. 1–34, 2023.

[3] Y. Yang et al., "Fuzzy-based deep attributed graph clustering," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 4, pp. 1951–1964, Apr. 2024.

[4] H. J. Li, Y. Feng, C. Xia, and J. Cao, "Overlapping graph clustering in attributed networks via generalized cluster potential game," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 1, pp. 1–26, 2024.

[5] K. Berahmand, M. Mohammadi, F. Saberi-Movahed, Y. Li, and Y. Xu, "Graph regularized nonnegative matrix factorization for community detection in attributed networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 1, pp. 372–385, Jan./Feb. 2023.

[6] Z. Bu, H. -J. Li, C. Zhang, J. Cao, A. Li, and Y. Shi, "Graph K-means based on leader identification, dynamic game and opinion dynamics," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 7, pp. 1348–1361, Jul. 2020.

[7] D. Jin et al., "A survey of community detection approaches: From statistical modeling to deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 1149–1170, Feb. 2023.

[8] H. Van Lierde, T. W. S. Chow, and G. Chen, "Scalable spectral clustering for overlapping community detection in large-scale networks," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 4, pp. 754–767, Apr. 2020.

[9] V. D. Blondel, J. -L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mechanics, Theory Exp.*, vol. 2008, no. 10, 2008, Art. no. P10008.

[10] Z. Bu, J. Cao, H. -J. Li, G. Gao, and H. Tao, "GLEAM: A graph clustering framework based on potential game optimization for large-scale social networks," *Knowl. Inf. Syst.*, vol. 55, no. 3, pp. 741–770, 2018.

[11] S. Fortunato and M. Barthelemy, "Resolution limit in community detection," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 1, pp. 36–41, 2007.

[12] H. Li, H. Cao, Y. Feng, X. Li, and J. Pei, "Optimization of graph clustering inspired by dynamic belief systems," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, pp. 6773–6785, Nov. 2024.

[13] A. Prat-Perez, D. Dominguez-Sal, and J. -L. Larriba-Pey, "High quality, scalable and parallel community detection for large real graphs," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 225–236.

[14] W. Zhen, C. Wang, C. Gao, X. Li, and X. Li, "An evolutionary autoencoder for dynamic community detection," *Sci. China, Inf. Sci.*, vol. 63, no. 11, 2023, Art. no. 212205.

[15] X. Li, S. Yin, X. Liu, C. Gao, Z. Wang, and V. I. Nekorkin, "Consensus subspace graph regularization based on prior information for multiplex network clustering," *Eng. Appl. Artif. Intell.*, vol. 135, 2024, Art. no. 108851.

[16] K. Deng, P. G. Mehta, and S. P. Meyn, "Optimal Kullback-Leibler aggregation via spectral theory of Markov chains," *IEEE Trans. Autom. Control*, vol. 56, no. 12, pp. 2793–2808, Dec. 2011.
[17] Z. Bu, H. -J. Li, J. Cao, Z. Wang, and G. Gao, "Dynamic cluster formation game for attributed graph clustering," *IEEE Trans. Cybern.*, vol. 49, no. 1, pp. 328–341, Jan. 2019.
[18] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 4, pp. 1118–1123, 2008.
[19] P. Pons and M. Latapy, "Computing communities in large networks using random walks," *J. Graph Algorithms Appl.*, vol. 10, no. 2, pp. 191–218, 2006.
[20] G. Karypis and V. Kumar, "Multilevel K-way hypergraph partitioning," *VLSI Des.*, vol. 11, no. 3, pp. 285–300, 2000.
[21] G. Karypis, E. -H. Han, and V. Kumar, "Chameleon: Hierarchical clustering using dynamic modeling," *Computer*, vol. 32, no. 8, pp. 68–75, Aug. 1999.
[22] J. Yang and J. Leskovec, "Overlapping community detection at scale: A nonnegative matrix factorization approach," in *Proc. 6th ACM Int. Conf. Web Search Data Mining*, 2013, pp. 587–596.
[23] E. Akbas and P. Zhao, "Attributed graph clustering: An attribute-aware graph embedding approach," in *Proc. 2017 IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, 2017, pp. 305–308.
[24] Z. Wang, C. Wang, X. Li, C. Gao, X. Li, and J. Zhu, "Evolutionary markov dynamics for network community detection," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 3, pp. 1206–1220, Mar. 2022.
[25] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *Proc. IEEE 13th Int. Conf. Data Mining*, 2013, pp. 1151–1156.
[26] S. Gunnemann, B. Boden, I. Farber, and T. Seidl, "Efficient mining of combined subspace and subgraph clusters in graphs with feature vectors," in *Proc. 17th Pacific-Asia Conf. Adv. Knowl. Discov. Data Mining*, Gold Coast, QLD, Australia, 2013, pp. 261–275.
[27] H. Cheng, Y. Zhou, and J. X. Yu, "Clustering large attributed graphs: A balance between structural and attribute similarities," *ACM Trans. Knowl. Discov. Data*, vol. 5, no. 2, 2011, Art. no. 12.
[28] Z. Li, J. Liu, and K. Wu, "A multiobjective evolutionary algorithm based on structural and attribute similarities for community detection in attributed networks," *IEEE Trans. Cybern.*, vol. 48, no. 7, pp. 1963–1976, Jul. 2018.
[29] Y. Zheng, C. Jia, and J. Yu, "Attributed graph clustering under the contrastive mechanism with cluster-preserving augmentation," *Inf. Sci.*, vol. 681, 2024, Art. no. 121225.
[30] X. Li, H. Zhang, R. Wang, and F. Nie, "Multiview clustering: A scalable and parameter-free bipartite graph fusion method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 330–344, Jan. 2022.
[31] C. Fang and Z. Z. Lin, "Overlapping communities detection based on cluster-ability optimization," *Neurocomputing*, vol. 494, pp. 336–345, 2022.
[32] B. Yang, J. Liu, and D. Liu, "Characterizing and extracting multiplex patterns in complex networks," *IEEE Trans. Syst. Man Cybern.-Part B (Cybern.)*, vol. 42, no. 2, pp. 469–481, Apr. 2012.
[33] A. Marin and S. Rossi, "On the relations between Markov chain lumpability and reversibility," *Acta Informatica*, vol. 54, no. 5, pp. 447–485, 2017.
[34] Z. Yang, J. Cor, and E. Oja, "Low-rank doubly stochastic matrix decomposition for cluster analysis," *J. Mach. Learn. Res.*, vol. 17, no. 187, pp. 1–25, 2016.
[35] E. Ravasz and A. L. Barabasi, "Hierarchical organization in complex networks," *Phys. Rev. E*, vol. 67, no. 2, 2003, Art. no. 026112.
[36] A. Arenas, A. Fernandez, and S. Gomez, "Analysis of the structure of complex networks at different resolution levels," *New J. Phys.*, vol. 10, no. 5, 2008, Art. no. 053039.

Huijia Li received the Ph.D. degree in operational research and control theory from the Academy of Mathematics & Systems Science, Chinese Academy of Sciences, Beijing, China, in 2013. He is currently a Professor with the School of Statistics and Data Science, Nankai University, Tianjin, China. His current research interests include data mining, machine learning, statistical analysis, and artificial intelligence.

Fanghao Lou received the B.S. degree in data science and Big Data technology from Nankai University, where he is currently working toward the M. S. degree in the field of statistics with the School of Statistics and Data Science. His current research interests include deep learning, graph neural networks, and data mining.

Qiqi Wang received the Ph.D. degree in natural language processing and LegalAI from the University of Auckland, Auckland, New Zealand, in 2024. He is currently a Lecturer with the School of Statistics and Data Science, Nankai University, Tianjin, China. His research interests include text mining, natural language processing, and artificial intelligence.

Guijun Li received the Ph.D. degree in management science from the Harbin Institute of Technology, Harbin, China, in 2003. He is currently a Full Professor with the School of Mathematical Sciences and also the Chancellor of Hebei Normal University, Hijiazhuang, China. He has authored or coauthored more than 100 high quality papers. His current research interests include data mining, management science, machine learning, and artificial intelligence.

Matjaž Perc received the Ph.D. degree in physics from the University of Maribor, Maribor, Slovenia, in 2006. He is currently a Professor of physics with the University of Maribor, Staff Researcher with Community Healthcare Center Dr. Adolf Drolc Maribor, Adjunct Professor with Kyung Hee University, Seoul, South Korea and Korea University, Seoul, and External Faculty Member with Complexity Science Hub in Vienna. He is a member of Academia Europaea and the European Academy of Sciences and Arts, and among top 1% most cited physicists according to 2020, 2021, 2022, 2023, and 2024 Clarivate Analytics data. He was the recipient of the 2015 Young Scientist Award for Socio and Econophysics from the German Physical Society, 2017 USERN Laureate, and the Zois Award, which is the highest national research award in Slovenia, in 2018. In 2019, he became a fellow of the American Physical Society. Since 2021, he has been the Vice-Dean of natural sciences at the European Academy of Sciences and Arts.