# METHODS FOR CLUSTERING MULTI-LAYER GRAPHS IN MOBILE NETWORKS

*Xiaowen Dong* †, *Pascal Frossard* †, *Pierre Vandergheynst* † *and Nikolai Nefedov* ‡

† Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
‡ Nokia Research Center (NRC), Lausanne, Switzerland
{xiaowen.dong, pascal.frossard, pierre.vandergheynst}@epfl.ch, nikolai.nefedov@nokia.com

## 1. INTRODUCTION

Clustering on graphs has been studied extensively for years due to its numerous applications. However, in contrast to the classic problems, clustering in mobile and online social networks brings new challenges. In these scenarios, it is common that observational data contains multiple modalities of information reflecting different aspects of human behavior and social interactions. These interactions may be represented by a multi-layer graph that share the same set of vertices representing users, while having different layers representing different relationships among users. Intuitively, each graph should contribute to a better understanding of the underlying clusters from its own angle. It may be expected that a proper combination of the multiple graphs could lead to a better unified clustering of users' behavior and their social interactions.

In this work we consider different methods to combine multi-layer graphs. In particular, we propose an efficient way to combine spectra of multiple graphs to form a "common spectrum". To verify the suggested approach we tested it using mobile datasets. Also we compare the proposed approach with community detection methods based on modularity maximization over single and multiple layer graphs.

## 2. GRAPH REGULARIZATION FRAMEWORK

The idea of working with the spectrum of the graph is inspired by the popular spectral clustering algorithm [1]. On a single graph, it applies eigen-decomposition of the graph Laplacian matrix and form a spectral embedding of the original vertices in a low dimensional space. This enhances the intrinsic relationship among vertices so that clustering based on this new representation is usually trivial. The problem is more complicated in case of multiple graph layers. As two recent examples, the authors of [2] use an unified matrix factorization framework to find a common low dimensional representation shared by the multiple graphs in the original space domain, while in [3] the authors propose a co-regularization framework to find such a representation in the graph spectral domain.

In this paper we generalize the one-layer spectral clustering to multiple graphs by finding a common low dimensional representation that captures the characteristics of all graph layers. More specifically, we propose first a graph regularization framework to combine the spectra of two graph layers. The key point is that we treat the eigenvectors of Laplacian matrix from one graph as functions defined on the vertices of another graph. By enforcing the "smoothness" of such functions on the second graph through a regularization framework, we capture the characteristics of both graphs and get a better unified clustering result than using single graphs separately. Moreover, our approach has several interpretations: it can be viewed as a propagation process of the cluster labels on the graph, as well as a framework to minimize a mismatch between the resulting partition and information from each individual graph. Next, we generalize this process to the case which involves more than two graphs.

## 3. MULTI-RESOLUTION COMMUNITIES DETECTION

To evaluate performance of the suggested approach above we compare it with modularity maximization [4] using fast greedy search algorithm [5]. Note that modularity maximization may give a different number of communities at different layers. On the other hand,

**Table 1**. MIT datasets: combination of phone-calls, BT and location layers. Evaluation of clustering performance using the proposed and the baseline methods. NMI and RI stand for normalized mutual information Rand index.

| | NMI | Purity | RI |
|---|---|---|---|
| The proposed method | 0.518 | 0.712 | 0.758 |
| Sum of spectral kernels | 0.486 | 0.673 | 0.729 |
| Sum of norm. adj. matrices | 0.484 | 0.685 | 0.753 |
| Sum of adj. matrices | 0.366 | 0.641 | 0.731 |

**Table 2**. NRC datasets: combination of phone-calls, BT and GPS layers. Evaluation of clustering performance using the proposed and the baseline methods.

| | NMI | Purity | RI |
|---|---|---|---|
| The proposed method | 0.395 | 0.539 | 0.708 |
| Community detection $^{(*)}$ | 0.363 | 0.507 | 0.628 |
| Sum of norm. adj. matrices | 0.381 | 0.534 | 0.710 |
| Sum of adj. matrices | 0.278 | 0.475 | 0.650 |
| Sum of spectral kernels | 0.220 | 0.378 | 0.570 |

$^{(*)}$ graph formed by summation of adjacency matrices.

the ground truth data typically is clustered into a fixed number of groups. To obtain the same number of communities at different layers as in the ground truth data we apply random walk approach [6].

In general, the suggested framework of a "common spectrum" may be implemented using the community detection approach (to appear elsewhere).

## 4. APPLICATIONS TO MOBILE DATASETS

We evaluate performance of the proposed clustering methods on the mobile phone datasets collected by MIT Media Lab [7] and Nokia Research Center (NRC) Lausanne [8]. In particular, we consider graph layers formed by phone-calls, detected WLAN and bluetooth proximity, and GPS locations. Simulations show that our approach to combine graph layers improves reliability of clustering compared to a several base-line methods [2] (see Table 1 and Table 2).

Furthermore, the concept of a "common spectrum" is helpful in analysis of any multimodal data which can be conveniently modeled as multiple graphs. For instance, it would enable us to generalize the normal spectral analysis from one-dimensional to multi-dimensional cases.

## 5. REFERENCES


[1] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug 2000.

[2] W. Tang, Z. Lu, and I. Dhillon, "Clustering with Multiple Graphs," in *International Conference on Data Mining*, Miami, Florida, USA, Dec 2009.

[3] Abhishek Kumar, Piyush Rai, and Hal Daumé III, "Co-regularized Spectral Clustering with Multiple Kernels," in *NIPS Workshop: New Directions in Multiple Kernel Learning*, 2010.

[4] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, vol. 69, no. 066133, 2004.

[5] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and Etienne Lefebvre, "Fast unfolding of communites in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 1742-5468, no. 10, pp. P10008+12, 2008.

[6] R. Lambiotte, J.-C. Delvenne, and M. Barahona, "Laplacian dynamics and multiscale modular structure in networks," *arXiv:0812.1770v3*, 2009.

[7] N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure Using Mobile Phone Data," in *Proceedings of the National Academy of Sciences*, 2009, vol. 106, pp. 15274–15278.

[8] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila, "Towards Rich Mobile Phone Datasets: Lausanne Data Collection Campaign," in *International Conference on Pervasive Services*, Berlin, Germany, Jul 2010.