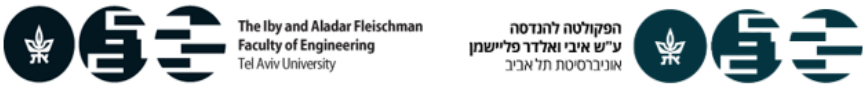

COMMUNITY DETECTION WITH APPLICATIONS TO MULTIREFERENCE ALIGNMENT

Project Number: 20.1.1.2135

Final report

by

Tomer Matityahu 312116668

Alexey Kiryushkin 324439876

Mentors:

Mr. Noam Janco Tel-Aviv University

Dr. Tamir Bendory Tel-Aviv University

Presented to the

School of Electrical Engineering

within the Tel-Aviv University

The Iby and Aladar Fleischman Faculty of Engineering

in partial fulfillment of

the requirements for the degree of

B.Sc Electrical Engineering

Project was done from home

TABLE OF CONTENTS

ABSTRACT

Single-particle reconstruction in Cryogenic Electron Microscopy (cryo-EM)[1] is a tool for constructing a 3D model of a biological macromolecule using 2D projections of the macromolecules taken by an electron microscope. An unsupervised classification of the 2D images is required in order to separate macromolecular projections of different conformations. Due to high noise levels and data heterogeneity, sophisticated clustering methods are needed.
In our project we will use Community Detection (CD) algorithms to cluster data generated from the Multireference Alignment (MRA) statistical model. The model abstracts away much of the intricacy of cryo-EM while retaining some of its essential features. Conclusions should be added

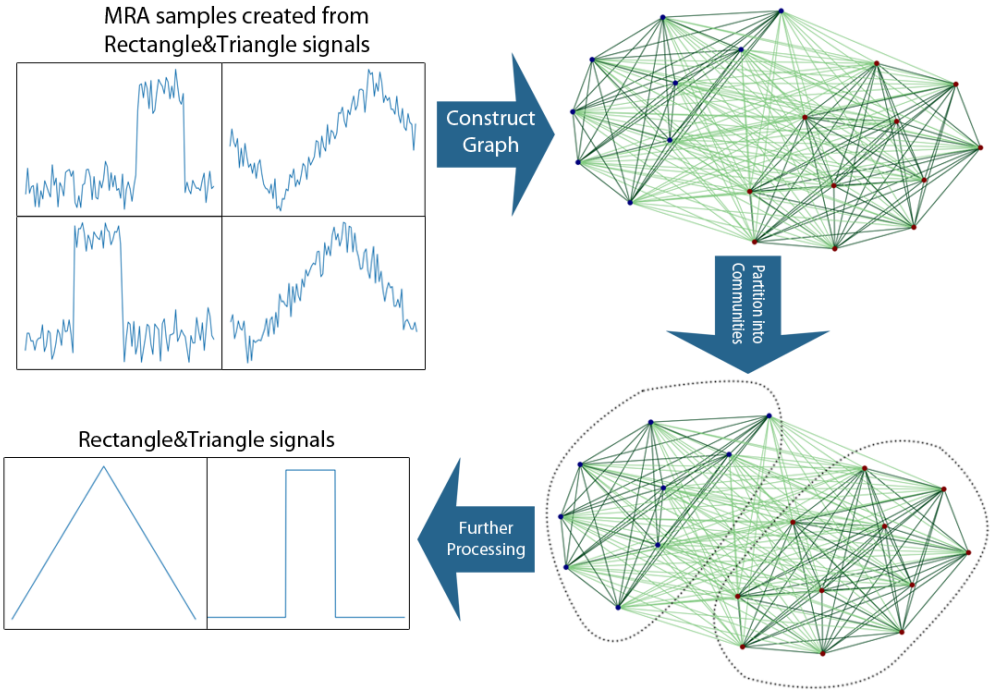

Figure 1: Project process. Further processing stage presents the idea behind clustering the data and is outside of the scope of the project.

List of Figures

List of Tables

List of Abbreviations

C

CD Community Detection. 3, 7

cryo-EM Cryogenic Electron Microscopy. 3, 7, 8

F

FFT Fast Fourier Transform. 8

M

MRA Multireference Alignment. 1, 3, 4, 7, 8

S

SNR Signal To Noise Ratio. 4, 7, 8

1 Introduction

Single-particle reconstruction in cryo-EM is a powerful image-processing tool used to determine the 3D structure of biological macromolecular complexes. 2D images (micrographs) of a macromolecule are taken by an electron microscope. Essentially, the set of all micrographs for a given macromolecule spans a 3D model of the macromolecule. Thus, single-particle reconstruction is using the micrographs to build a 3D model of the macromolecule.
Due to high sensitivity of the biological macromolecules to radiation damage, electron microscope provides limited electron doses when producing micrographs. This and the low contrast of micrographs result in cryo-EM data having very low Signal To Noise Ratio (SNR)[1].
cryo-EM technology has the potential to offer the ability to analyze different functional and conformational states of macromolecules, an important ability for the field of molecular biology. Practically, it entails the classification of heterogeneous cryo-EM data.
Many different approaches for cryo-EM data classification have been developed. Typically, likelihood optimization algorithms and Bayesian inference frameworks are used to deal with data heterogeneity[6, 5, 4, 7, 3].
In our project we propose a different approach to cryo-EM data classification using Community Detection (CD) algorithms from the field of complex networks. Namely, data will be classified following the steps:

- Converting data into a graph
- Applying Community Detection algorithms to partition the graph into distinct communities
- Each vertice in a graph that corresponds with a single sample will be given a label

For the sake of an abstraction of the cryo-EM data we will use the Heterogeneous Multireference Alignment (MRA) statistical model. Throughout our project we use the simplified 1D version of the model.

2 Background

2.1 Heterogeneous 1D Multireference Alignment model

In our project we will use the Heterogeneous 1D Multireference Alignment statistical model[2] (MRA shortly) for the sake of cryo-EM data abstraction. Bellow is the definition of the model.
Let $x_1, ..., x_K \in \mathbb{R}^L$ be K unknown normalized signals (distinct even up to shift) and let R_s be the cyclic shift operator: $(R_s x)[n] = x[\langle n - s \rangle_L]$. We are given N observations:

$$y_j = R_{s_j} x_{k_j} + \varepsilon_j, \quad j = 1, ..., N \tag{1}$$

where $s_j \sim U[0, L-1], k_j \sim U[0, K-1]$ and $\varepsilon_j \sim \mathcal{N}(0, \sigma^2 I)$ is i.i.d white Gaussian noise. Our goal is to estimate the signals $x_1, ..., x_K$ from the observations.
Simply speaking, MRA observation is a randomly chosen signal x_{k_j}, shifted randomly by s_j and distorted using white noise. **Figure 2** shows an example of two MRA observations at different noise levels.

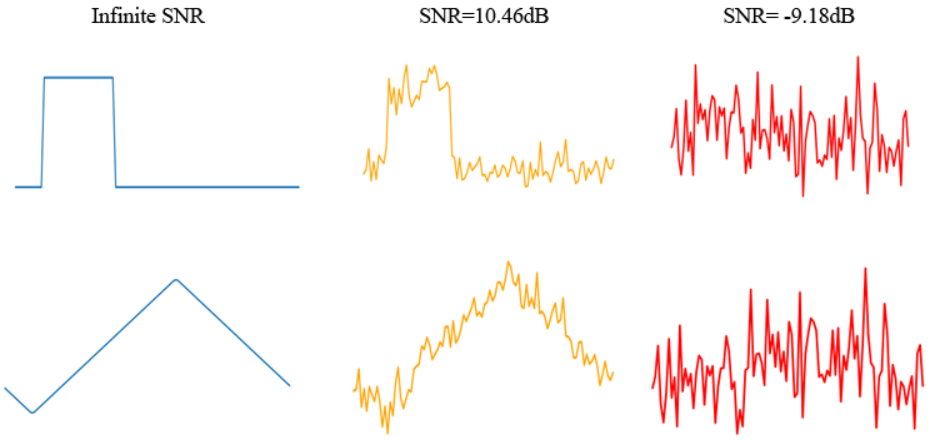

Figure 2: Example of MRA observations at different SNR levels. Each column shows a shifted distinct signal at different noise level. We can see that for low SNR the task of signal estimation is quite challenging.

In order to build a graph from MRA observations, a similarity measure between observations must be defined. We will use the cross-correlation, as its invariance to shift holds great value.

$$(x \star y)_n \triangleq \sum_{l=\infty}^{\infty} x_l^*(y)_{n+l} \tag{2}$$

The convolution theorem states that the convolution of two signals equals to the inverse Fourier transform of the product of the Fourier transforms of each signal. Thus we can write the equation above in terms of Fourier transforms and later exploit FFT in our simulations.

$$(x \star y)_n = \mathcal{F}^{-1}\{X^* \cdot Y\}_n \tag{3}$$

In order to obtain normalized cross-correlation MRA samples must first be normalized.

2.2 Introduction to graphs

3 Aim 1 Title

3.1 Introduction to Aim 1

An introduction to Aim 1.

3.2 Background to Aim 1

This section will include the most relevant literature addressing this aim.

3.3 Methods

Maybe you'll discuss some methods.

3.3.1 Some crucial details about the method

It'll probably have a sub(sub)heading.

3.3.2 Conceptual model, research questions and hypotheses

Blah blah blah.

3.4 Results of Aim 1

Blah blah blah.

3.5 Discussion of Aim 1

Blah blah blah.

3.6 Conclusion of Aim 1

Blah blah blah.

4 Aim 2 Title

4.1 Introduction to Aim 2

An introduction to Aim 2.

4.2 Background to Aim 2

This section will include the most relevant literature addressing this aim.

4.3 Methods

Maybe you'll discuss some methods.

4.3.1 Some crucial details about the method

It'll probably have a sub(sub)heading.

4.3.2 Conceptual model, research questions and hypotheses

Blah blah blah.

4.4 Results of Aim 2

Blah blah blah.

4.5 Discussion of Aim 2

Blah blah blah.

4.6 Conclusion of Aim 2

Blah blah blah.

5 Aim 3 Title

5.1 Introduction to Aim 3

An introduction to Aim 3.

5.2 Background to Aim 3

This section will include the most relevant literature addressing this aim.

5.3 Methods

Maybe you'll discuss some methods.

5.3.1 Some crucial details about the method

It'll probably have a sub(sub)heading.

5.3.2 Conceptual model, research questions and hypotheses

Blah blah blah.

5.4 Results of Aim 3

Blah blah blah.

5.5 Discussion of Aim 3

Blah blah blah.

5.6 Conclusion of Aim 3

Blah blah blah.

6 Discussion

Some detailed discussion.

6.1 A subheading

Blah blah blah

7 Conclusion

This section would contain the conclusions drawn from the entire body of work.

7.1 A subheading

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

8 References

[1] Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Processing Magazine*, 37(2):58–76, 2020.

[2] Nicolas Boumal, Tamir Bendory, Roy R. Lederman, and Amit Singer. Heterogeneous multireference alignment: a single pass approach, 2018.

[3] Saikat Chowdhury, Stephanie A Ketcham, Trina A Schroer, and Gabriel C Lander. Structural organization of the dynein–dynactin complex bound to microtubules. *Nature structural & molecular biology*, 22(4):345–347, 2015.

[4] Sjors HW Scheres. Beam-induced motion correction for sub-megadalton cryo-em particles. *elife*, 3:e03665, 2014.

[5] Sjors HW Scheres, Mikel Valle, and José-María Carazo. Fast maximum-likelihood refinement of electron microscopy images. *Bioinformatics*, 21(suppl_2):ii243–ii244, 2005.

[6] Fred J Sigworth. A maximum-likelihood approach to single-particle image refinement. *Journal of structural biology*, 122(3):328–339, 1998.

[7] Chun Feng Song, Kostas Papachristos, Shaun Rawson, Markus Huss, Helmut Wieczorek, Emanuele Paci, John Trinick, Michael A Harrison, and Stephen P Muench. Flexibility within the rotor and stators of the vacuolar h+-atpase. *PLoS One*, 8(12):e82207, 2013.

9 Appendix

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.