

UNIVERSITATEA "ALEXANDRU-IOAN CUZA" DIN IAȘI

FACULTATEA DE INFORMATICĂ



LUCRARE DE LICENȚĂ

Aspecte computaționale în biologie

propusă de

Andrei-Ionuț Popescu

Sesiunea: februarie, 2019

Coordonator științific

Conf. Dr. Ionescu Ionel

UNIVERSITATEA "ALEXANDRU-IOAN CUZA" DIN IAȘI

FACULTATEA DE INFORMATICĂ

Aspecte computaționale în biologie

Andrei-Ionuț Popescu

Sesiunea: februarie, 2019

Coordonator științific

Conf. Dr. Ionescu Ionel

Avizat,
Îndrumător lucrare de licență,
Conf. Dr. Ionescu Ionel.

Data: Semnătura:

Declarație privind originalitatea conținutului lucrării de licență

Subsemnatul **Popescu Andrei-Ionuț** domiciliat în **România, jud. Iași, mun. Iași, calea Buzăului, nr. 25, bl. A, et. 5, ap. 45**, născut la data de **01 ianuarie 2018**, identificat prin CNP **1234567891234**, absolvent al Facultății de informatică, **Facultatea de informatică** specializarea **informatică**, promoția 2018, declar pe propria răspundere cunoscând consecințele falsului în declarații în sensul art. 326 din Noul Cod Penal și dispozițiile Legii Educației Naționale nr. 1/2011 art. 143 al. 4 și 5 referitoare la plagiat, că lucrarea de licență cu titlul **Aspecte computaționale în biologie** elaborată sub îndrumarea domnului **Conf. Dr. Ionescu Ionel**, pe care urmează să o susțin în fața comisiei este originală, îmi aparține și îmi asum conținutul său în întregime.

De asemenea, declar că sunt de acord ca lucrarea mea de licență să fie verificată prin orice modalitate legală pentru confirmarea originalității, consimțind inclusiv la introducerea conținutului ei într-o bază de date în acest scop.

Am luat la cunoștință despre faptul că este interzisă comercializarea de lucrări științifice în vederea facilitării falsificării de către cumpărător a calității de autor al unei lucrări de licență, de diplomă sau de disertație și în acest sens, declar pe proprie răspundere că lucrarea de față nu a fost copiată ci reprezintă rodul cercetării pe care am întreprins-o.

Data:

Semnătura:

Declarație de consimțământ

Prin prezenta declar că sunt de acord ca lucrarea de licență cu titlul **Aspecte computaționale în biologie**, codul sursă al programelor și celelalte conținuturi (grafice, multimedia, date de test, etc.) care însoțesc această lucrare să fie utilizate în cadrul Facultății de informatică.

De asemenea, sunt de acord ca Facultatea de informatică de la Universitatea "Alexandru-Ioan Cuza" din Iași, să utilizeze, modifice, reproducă și să distribuie în scopuri necomerciale programele-calculator, format executabil și sursă, realizate de mine în cadrul prezentei lucrări de licență.

Absolvent **Andrei-Ionuț Popescu**

Data:

Semnătura:

Cuprins

Motivație	2
Introducere	3
1 Description of methods	4
1.1 Graph Clustering	5
1.1.1 Short intro about what graph clustering is	5
1.1.2 Why graph clustering instead other traditional methods such as k-means, density based techniques etc	5
1.1.3 Types of graph clustering	5
1.1.4 Community detection - Optimizing the quality function	5
1.1.5 Louvain	5
1.1.6 Louvain refined	5
1.1.7 SLM	5
1.1.8 Leiden	5
1.2 PhenoGraph pipeline	5
1.2.1 Describing the pipeline	5
1.2.2 How to convert matrix data into a graph using kNN	5
1.2.3 SNN - providing weights using Jaccard Similarity Index	5
1.3 Element-Centric Similarity	5
1.3.1 Description about how it works	5
1.3.2 Properties, comparison with other clustering metrics	6
1.3.3 ECC	6
1.4 Intro info about biological data and sequencing techniques	6
2 The importance of parameter values in the clustering output	7
2.1 Monocle and Seurat	7

3	ClustAssess	8
3.1	Titlul secțiunii 1	8
3.2	Titlul secțiunii 2	9
4	Experiments and results	10
	Conclusions, Future Work	11

Motivație

Diam sit amet nisl suscipit adipiscing bibendum. Aliquet lectus proin nibh nisl condimentum id. Urna duis convallis convallis tellus id interdum velit laoreet. Amet tellus cras adipiscing enim eu turpis egestas pretium aenean. Tortor condimentum lacinia quis vel eros donec ac odio tempor. Volutpat ac tincidunt vitae semper. Urna cursus eget nunc scelerisque viverra mauris in aliquam. Aliquam id diam maecenas ultricies. Molestie a iaculis at erat. Tincidunt nunc pulvinar sapien et ligula ullamcorper malesuada proin. Consequat interdum varius sit amet. Eget est lorem ipsum dolor sit amet consectetur adipiscing. Pharetra diam sit amet nisl suscipit adipiscing bibendum. Maecenas sed enim ut sem viverra aliquet eget sit. Enim blandit volutpat maecenas volutpat blandit aliquam etiam erat velit.

Introducere

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Nunc mattis enim ut tellus elementum sagittis vitae et. Placerat in egestas erat imperdiet sed euismod. Urna id volutpat lacus laoreet non curabitur gravida. Blandit turpis cursus in hac habitasse platea. Eget nunc lobortis mattis aliquam faucibus. Est pellentesque elit ullamcorper dignissim cras tincidunt lobortis feugiat. Viverra maecenas accumsan lacus vel facilisis volutpat est. Non odio euismod lacinia at quis risus sed vulputate odio. Consequat ac felis donec et odio pellentesque diam volutpat commodo. Etiam sit amet nisl purus in. Tortor condimentum lacinia quis vel eros donec. Phasellus egestas tellus rutrum tellus pellentesque eu tincidunt. Aliquam id diam maecenas ultricies mi eget mauris pharetra. Enim eu turpis egestas pretium.

Capitolul 1

Description of methods

This chapter contains informations about the methods used for graph clustering, the sequencing and processing the biological data and eventually mentions of other works / papers that were focusing on assessing the robustness on changing the seed.

1.1 Graph Clustering

1.1.1 Short intro about what graph clustering is

1.1.2 Why graph clustering instead other traditional methods such as k-means, density based techniques etc

1.1.3 Types of graph clustering

1.1.4 Community detection - Optimizing the quality function

1.1.5 Louvain

1.1.6 Louvain refined

1.1.7 SLM

1.1.8 Leiden

1.2 PhenoGraph pipeline

1.2.1 Describing the pipeline

Present the steps that describe the pipeline. (Dimensionality reduction, graph building and graph clustering)

About dimensionality reduction

1.2.2 How to convert matrix data into a graph using kNN

1.2.3 SNN - providing weights using Jaccard Similarity Index

1.3 Element-Centric Similarity

1.3.1 Description about how it works

Describe the intuition behind ECS: the idea of the bipartite graph between points and clusters.

More details about how to calculate ECS. Talk about the affinity matrix and the L1 distance.

1.3.2 Properties, comparison with other clustering metrics

Present some limitation of other clustering metrics such as bias toward cluster sizes, shapes and so on. Perhaps present some comparison figures from the main article.

Present some properties of ECS:

1. the fact that it can be used not only for flat disjoint clusterings, but also for overlapping or hierarchical partitions
2. it overcomes the biases present in the other clustering metrics
3. ECS illustrates the overall similarity between two partitions but also can help in identifying the points where the clustering are not similar

1.3.3 ECC

Talk about how ECC is calculated

1.4 Intro info about biological data and sequencing techniques

Tell about sequencing techniques, how the initial data looks, about cells, genes, what they mean, what is the role and the purpose of the clusters in the biological interpretation.

Capitolul 2

The importance of parameter values in the clustering output

2.1 Monocle and Seurat

Introduce some details about these packages, the language

Capitolul 3

ClustAssess

Amet venenatis urna cursus eget. Quam vulputate dignissim suspendisse in est ante. Proin nibh nisl condimentum id. Egestas maecenas pharetra convallis posuere morbi. Risus viverra adipiscing at in. Vulputate eu scelerisque felis imperdiet. Cras adipiscing enim eu turpis egestas pretium aenean pharetra. In aliquam sem fringilla ut morbi tincidunt augue. Montes nascetur ridiculus mus mauris. Viverra accumsan in nisl nisi scelerisque eu ultrices vitae. In nibh mauris cursus mattis molestie a iaculis. Interdum consectetur libero id faucibus nisl tincidunt eget. Gravida in fermentum et sollicitudin ac orci. Suscipit adipiscing bibendum est ultricies. Etiam non quam lacus suspendisse. Leo urna molestie at elementum eu facilisis sed odio morbi. Egestas congue quisque egestas diam in arcu cursus. Amet consectetur adipiscing elit ut aliquam purus.

3.1 Titlul secțiunii 1

Eros donec ac odio tempor. Facilisi morbi tempus iaculis urna id volutpat. Faucibus in ornare quam viverra orci sagittis eu. Amet tellus cras adipiscing enim eu turpis egestas. Integer feugiat scelerisque varius morbi. Platea dictumst vestibulum rhoncus est pellentesque elit ullamcorper dignissim. Bibendum arcu vitae elementum curabitur. Eu nisl nunc mi ipsum faucibus. Id aliquet lectus proin nibh nisl condimentum id venenatis a. Cras adipiscing enim eu turpis egestas pretium. Quisque non tellus orci ac auctor augue mauris augue. Malesuada pellentesque elit eget gravida cum. Ut lectus arcu bibendum at. Massa id neque aliquam vestibulum morbi blandit. Posuere ac ut consequat semper viverra nam. Viverra adipiscing at in tellus integer feugiat

scelerisque varius morbi. Morbi enim nunc faucibus a pellentesque sit amet porttitor eget. Eu feugiat pretium nibh ipsum consequat nisl vel. Nisl purus in mollis nunc sed.

3.2 Titlul secțiunii 2

Elementum sagittis vitae et leo duis ut diam quam nulla. Purus sit amet volutpat consequat mauris nunc. Tincidunt augue interdum velit euismod in pellentesque massa. Nunc sed augue lacus viverra vitae congue. Porttitor leo a diam sollicitudin. Faucibus pulvinar elementum integer enim. Adipiscing bibendum est ultricies integer quis auctor elit. Blandit aliquam etiam erat velit scelerisque in. A iaculis at erat pellentesque adipiscing commodo elit at. Erat nam at lectus urna duis. Consequat ac felis donec et. Fermentum posuere urna nec tincidunt praesent semper feugiat nibh sed. Proin gravida hendrerit lectus a. Pretium viverra suspendisse potenti nullam ac tortor vitae purus. Arcu cursus euismod quis viverra nibh cras pulvinar mattis. Gravida arcu ac tortor dignissim convallis aenean. Quam nulla porttitor massa id neque aliquam vestibulum morbi. Sed viverra ipsum nunc aliquet. Quis enim lobortis scelerisque fermentum dui faucibus in.

Capitolul 4

Experiments and results

Conclusions, Future Work

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Nunc mattis enim ut tellus elementum sagittis vitae et. Placerat in egestas erat imperdiet sed euismod. Urna id volutpat lacus laoreet non curabitur gravida. Blandit turpis cursus in hac habitasse platea. Eget nunc lobortis mattis aliquam faucibus. Est pellentesque elit ullamcorper dignissim cras tincidunt lobortis feugiat. Viverra maecenas accumsan lacus vel facilisis volutpat est. Non odio euismod lacinia at quis risus sed vulputate odio. Consequat ac felis donec et odio pellentesque diam volutpat commodo. Etiam sit amet nisl purus in. Tortor condimentum lacinia quis vel eros donec. Phasellus egestas tellus rutrum tellus pellentesque eu tincidunt. Aliquam id diam maecenas ultricies mi eget mauris pharetra. Enim eu turpis egestas pretium.