Dissertation Proposal

Proposed Topic:

A taxonomy of group pattern discovery algorithms from spatio-temporal trajectories

xxx 21xxxxxxxx
Supervisor:
University of West London

Executive Summary

Over the years, the increasing development of location acquisition devices have generated a significant amount of spatio-temporal data. This data can be further analysed in search for some interesting patterns, new information, or to construct predictive models such as next location prediction. While data mining has been practiced for many years, the spatio-temporal data mining is a relatively new discipline and therefore requires certain standardization of the processes involved. The goal of this proposed dissertation is to contribute to the future research and development of group pattern discovery algorithms from spatio-temporal data by providing a conceptual framework for group pattern mining, followed by a comprehensive classification of existing models, their review and comparison. Additionally, since this dissertation is considered as the step towards developing a novel algorithm in the future, the result of this extensive literature review should bring the concept of an "ideal" model for group discovery from spatio- temporal trajectories. The framework will include both regular and big data processing models which to the best of author's knowledge is the first attempt of presenting both approaches in this context and including them in the data mining process. In addition to this, the current literature does not present any detailed publication on the classification of group pattern discovery algorithms from spatio-temporal trajectories.The proposed dissertation will take a form of a survey/taxonomy of the existing group discovery models from spatio-temporal trajectories.

1 Background of the Study

The growth of tracking technologies and development of location acquisition devices have generated big volumes of spatio-temporal data. This vast amount of data can be further analysed in order to bring valuable outcome by either discovering interesting findings and patterns, or by creating predictive models based on the given input. While data mining has been practiced for many years, the spatio-temporal data mining is a relatively new discipline and requires different approaches and techniques due to the nature of the input data. Spatio-temporal data is different from relational data as it adds both spatial and temporal attributes to regular observations. Given the example of a web search, in addition to the search request and related information, the spatio-temporal nature of the data would mean the stored request also comes with time and location details from where the search was made. Similarly, a CCTV, or other monitoring device would record the motion, or image of the given area together with its geo-coordinates and time the video, or image was taken. Shekhar et al. differentiates three distinct types of data attributes in the context of spatio-temporal data [Shekhar et al., 2015]. Aside from the spatial and temporal attributes, there are also non-spatiotemporal ones which provide additional characteristics of objects which could be for example their name (nominal), or exchange rate (ratio) among others. Spatial attributes can either define a location by providing for instance its geo-coordinates, spatial size, area or shape while temporal attributes can be anything from timestamps, snapshots, or duration. One of the challenges in spatio-temporal data mining process is dealing with spatio-temporal relationships. One of the approaches could be to transform them into a traditional data input so they can be modeled using traditional data mining techniques. Another - to find new ways of incorporating both spatial and temporal attributes into the data mining process.

In addition to the complexity of the spatio-temporal data itself, scalability comes as another challenge. With endless data generated every second heading towards zettabytes in the nearest future, the data mining algorithms need to cope with the volumes. Seeing the temporal aspect of such data and the increasing development of streaming technologies, the data mining process requires new solutions to speed up the processing. This could be for instance new data structures to store the relationship rather than the information which is too heavy to be processed (for e.g. [Tang et al., 2013]), or distributed/parallel processing models (for e.g.[Vatsavai et al., 2012, Orakzai et al., 2016]).

Spatio-temporal data types can be further classified basing on the way how space and time are used for data collection and representation. Atluri et al. specifies four categories of spatio-temporal data types [Atluri et al., 2017]. First of them is event data which includes discrete events occurring at point locations and times which can find its real-world application in crime detection. Another is raster data which can be best represented in the example for fMRI scans of brain activity. Point reference data is the third category with its application in surface temperature collection and lastly the trajectory data. Trajectory data is focused on measuring the trajectories of moving objects.

In his survey on trajectory data mining Zheng ([Zheng, 2015] classified four types of spatio-temporal data based on the moving objects. These types include trajectories of people, vehicles, animals and natural phenomena and each of them assuming various approaches of data mining. These trajectories have various applications including in intelligent transportation systems, social networks, or in commercial use (for e.g. to offer products based on the next location of the user). Given the example of transportation many taxis are equipped with GPS sensors to report on their locations to the transportation system or applications for the users to find their next available taxi. Scientists record trajectories of animals to analyse their behaviours and patterns for further research. Trajectory data mining has become an interdisciplinary field attracting specialists from many areas including computer science but also sociology, biology, or geography. Further in his survey on trajectory Zheng identifies four major categories of patterns from either single or group trajectories [Zheng, 2015]. These include trajectory clustering, sequential patterns, periodic patterns and moving together patterns. The first one aims at grouping the similar trajectories into patterns. The second looks at common sequences of locations in similar time intervals. The third one as the name suggests analyses the periodic occurrence of patterns. Lastly, the moving together patterns or in other words the group pattern mining from spatio-temporal trajectories will be the research area for this dissertation project which is further discussed in the following sections.

2 Literature Review

There have been several distinct approaches to group pattern mining using spatio-temporal data. One of the widely discussed approaches is flock pattern which is also considered one of the earliest group pattern discovery approaches [Benkert et al., 2008, Gudmundsson, 2004]. The second approach, convoy, argues that the first model is not sufficient for correct group discovery due to its limitation of pre-defined group search region which leads to group omissions (also known as so-called lossy-flock problem). Instead, this second approach proposes to capture trajectory pattern of any shape by applying the density-based clustering [Zheng, 2015, Jeung et al., 2008]. The third approach, swarm identifies another challenge with both aforementioned approaches as they have a strict requirement on the k-consecutive time points for group discovery [Li et al., 2010]. This may result in ignoring certain groups depending on the k-consecutive time set up. For instance if k is set up to 9 but the objects would travel together during k=7, then this group would have been missed. Instead, the swarm approach introduces a cluster of objects that can last for at least k but non-consecutive time stamps. While all these approaches use static datasets and additionally convoy and swarm require loading the entire trajectories into memories in order to identify the clusters, the traveling companion approach implements a data structure called *traveling buddy* [Tang et al., 2012] which allows finding patterns from streamed data and therefore sets up a ground for further approaches. Lastly, the most recent approach - loose traveling companion pattern addresses certain limitations of the previous approaches by focusing on human and not object/animal trajectories as presented in the previous patterns and considering indoor environment. A key observation is made that human trajectories are different to these of objects and animals as people can belong to one group although they have different trajectory paths (which is different to convoy) [Naserian et al., 2018].

The aforementioned approaches do not exhaust the group discovery patterns in spatio-temporal trajectories. Gathering, group pattern, moving cluster, evolving cluster, or weakly consistent group movement pattern are just few other examples that have contributed to the subject area. Furthermore, each of the classical approaches has its modifications such as variable flock which permits the members of the group to change during the time-span. Further analysis of these algorithms reveals certain commonalities they have as well as distinct features. Some of them have been summarized below:

PATTERN	AUTHORS	CONSECUTIVE/NONCONSECUTIVE TIME	GATHERING SHAPE	GROUP MEMBERSHIP	STATIC/STREAMING DATA
FLOCK	Benkert, Gudmudsson et al	k-consecutive	predefined (circular)	fixed	STATIC
CONVOY	Jeung et al	k-consecutive	any	fixed	STATIC
SWARM	Li et al	non-consecutive	any	fixed	STATIC
GATHERING	Zheng et al	non-consecutive	any but stable	evolving	STATIC
GROUP PATTERN	Wang et al	non-consecutive	predefined (circular)	fixed	STATIC
TRAVELLING COMPANION	Tang et al	k-consecutive	any	fixed	STREAMING
LOOSE COMPANION	Tang et al	non-consecutive	any	changeable	STREAMING
MOVING CLUSTER	Kalnis et al	k-consecutive	any/cluster	changeable	STATIC
EVOLVING CONVOY	Aung & Tan	k-consecutive	any/cluster	changeable	STATIC
WCM1	Wang et al	non-consecutive (time-gap threshold)	any/cluster	changeable	STATIC
LCTP2	Neserian et al	consecutive	any/cluster	changeable	STREAMING
WCLTCP3	Naserian et al	non-consecutive	any/cluster	changeable	STREAMING

Based on the reviewed literature the main features can be grouped around the continuity of the group (consecutive/non-consecutive time), how the group is formed (group formation), whether the members can change during the lifespan (group membership) and what data is used for pattern mining (static/streaming data). Additionally, each of the approaches can be seen from the scalability point of view and whether they can handle big data processing. Depending on each parameter and requirements there are significant differences in group pattern discovery which leads to several challenges in finding the best group pattern algorithm.

Due to the complexity of spatio-temporal data itself as well as its continuously growing and changing volume there are several challenges that are worth noticing and which have also been re-occurring in the related literature. These challenges can be summarized as per below:

- **Stream analysis**: Growth of streaming technologies which generates high volumes of data require a new approach to extract valuable information from such data
- **Incremental discovery**: Many current applications (especially monitoring or surveillance) require the output results simultaneously while receiving and processing the streamed data which means the algorithms should be able to discover group patterns in an incremental manner.
- **Efficiency**: Streaming means huge amount of data generated in a very short time which leads to extreme computational cost of processing. The algorithms should then be able to output the groups efficiently.
- **Big data processing**: Huge amounts of spatio-temporal trajectories data require a distributed or parallel approach to handling the volumes and to reduce the computational overload
- **Parameter-free algorithms**: Many algorithms include a large number of parameters and each of them has a significant impact on the performance and the precision of the results. There is no standardized way of choosing the right set of parameters rather than by conducting experiments which is not an ideal situation
- **Group continuity**: Traveling together may not always mean the members of the group stay together during the lifespan of the group. This is particularly noticeable in human trajectories and in larger groups where members of the group can form sub-groups [Naserian et al., 2018]
- **Effectiveness**: The number of members of the group can vary and the effective algorithm should be able to understand the wider picture and report long-term groups rather than short term ones.
- **High Precision**: The challenge lies with discovering the correct groups rather than more groups which can make a difference in certain applications based on personalized services and targeting the wrong group may be worse than not addressing it at all.
- **Data mining theory and formalization**: Esling as well as many other researchers noticed the lack of formalization of data mining theory such as through solid mathematical foundation [Esling and Agon, 2012, Zheng, 2015, Dodge et al., 2008].

Many researchers highlighted the importance of readjusting the currently existing algorithms to extract trajectories coming from streamed data and in the incremental manner rather than from static/historical datasets. This was one of the reasons why the traveling companion algorithm was presented together with a data structure *traveling buddy* which allowed storing the object relationships instead of their spatial coordinates [Tang et al., 2012]. Analyzing and mining streamed data comes at a very high computational cost and processing it in an incremental manner means that there is a huge amount of data flow at a very short time and many algorithms do not scale up to big data [Esling and Agon, 2012]. Therefore it is expected from the group discovery algorithms to be efficient [Naserian et al., 2018, Tang et al., 2012]. One of the ways of boosting the efficiency is to handle spatio-temporal trajectory data via distributed, or parallel computing. There are several examples of modifying the existing algorithms to handle big data processing. Orakzai et al. in their work [Orakzai et al., 2016], amend the convoy pattern for distributed processing while Vatsavai et al. analyze various algorithms from the distributed computing point of view [Vatsavai et al., 2012].

In his work on time-series data mining, Esling highlights another important research trend and challenge at the same time which is parameter-free data mining [Esling and Agon, 2012]. The majority (if not all of the reviewed algorithms) are hugely dependent on parameters/thresholds. These parameters affect both the performance of the algorithms and the precision of the results. Most of the discussed algorithms would include the size threshold and duration threshold (flock, convoy, swarm, traveling companion). Loose traveling companion includes additionally the frequency threshold and time-gap threshold. Given the example of the loose traveling companion setting up the size threshold higher results in a decrease of the time cost as it is easier to extract bigger groups. On the other hand, choosing the right value for the duration threshold depends on the type of the application and the other circumstances which can impact the number and precision of the identified groups. Taking as an example the closed locations with many changing groups of people such as airports or shopping centers setting up the duration threshold to a higher value such as days or weeks may result in completely different outcome to if it is set up to hours, or minutes. Some researchers attempted to create parameter-free models [Keogh et al., 2004] which are worth mentioning and ideally a perfect model should be parameter-free or have as few as possible so the results are not biased, nor in any way affected.

Specifically, in regards to group pattern discovery the challenge of defining a group has been taken by many researchers. Traveling together may not necessarily mean the members of the group will always stay together during their travel time and especially in human trajectories members of the group can split into sub-groups which require some flexibility in regards to how the continuity of the group is defined. One of the solutions as mentioned in the context of the parameters is to either modify the duration threshold or introduce a time-gap threshold allowing certain period of time when the members can stay apart. However, this is always subject to the right set up of the parameters which proves to be hard. Many state-of-art algorithms and clustering methods seem to ignore the fact that the group may split [Zheng et al., 2013, Li et al., 2010].

High precision and effectiveness of group discovery algorithms is strongly related with identifying the right groups rather than more groups and understanding the bigger picture. Similarly to the discovery of the sub-groups the correct groups identification is dependent on the right parameters [Naserian et al., 2018].

3 Statement of Problem

Taking into consideration all above challenges and the lack of a comprehensive classification of group pattern mining algorithms from spatio-temporal trajectories, there are several questions that can be answered in the proposed dissertation to fill that gap. The two of key research questions in this context are following:

1. What are the commonalities of the existing algorithms for group discovery from spatio-temporal trajectories?
2. Can they be generalized to form a conceptual framework for group pattern mining from spatio-temporal trajectories?

These can be further split into sub-questions to support the research in this area:

1. What are all currently available algorithms for group discovery from spatio-temporal trajectories?
2. What are their definitions?

3. What are their features?
4. What are their similarities and differences?
5. How they can be compared (complexity, scalability)?
6. What are currently known frameworks for group/data mining from spatio-temporal trajectories?
7. What are the challenges in this subject area?

4 Goal and Objectives

The goal of this proposed dissertation is to contribute to the future research and development of group pattern discovery algorithms from spatio-temporal data by providing a conceptual framework for group pattern mining, followed by a comprehensive classification of existing models, their review and comparison. Additionally, since this dissertation is considered as the step towards developing a novel algorithm in the future, the result of this extensive literature review should bring the concept of an "ideal" model for group discovery from spatio- temporal trajectories. The framework will include both regular and big data processing models which to the best of author's knowledge is the first attempt of presenting both approaches in this context. In addition to this the current literature does not present any detailed publication specifically on the classification of group pattern discovery algorithms from spatio-temporal trajectories. The closest to this concept are the works of Zheng [Zheng, 2015], Alturi [Atluri et al., 2017] and Dodge [Dodge et al., 2008]. Nevertheless, Zheng presents the whole trajectory data mining process where group pattern discovery or "moving together patterns" (as mentioned in that paper) is only briefly explained among three other trajectory pattern mining approaches as the final element in the discussed paradigm. Similarly, Alturi focused on the general spatio-temporal data mining process from the perspective of problems and methods while Dodge, whose work is the closest to this approach, provided a framework and classification of movement patterns from the perspective of the movements behaviors. Furthermore, the latter analyses all movement patterns including spatial, temporal as well as spatio-temporal solutions and was published in 2008 and therefore does not include the last 10 years of the research done in the spatio-temporal area which forms the basis of the proposed dissertation.

In order to achieve the goal of this dissertation project, it has been split into smaller objectives which are following:

- extracting definitions of the group pattern discovery algorithms from available literature
- extracting models of the algorithms from the available pseudo-codes and attached documentations
- grouping the definitions and models based on commonalities
- drafting a conceptual framework based on identified processes
- classifying the algorithms based on the framework
- performance review and comparison of the classified algorithms
- defining the "ideal" model features
- indications for future research

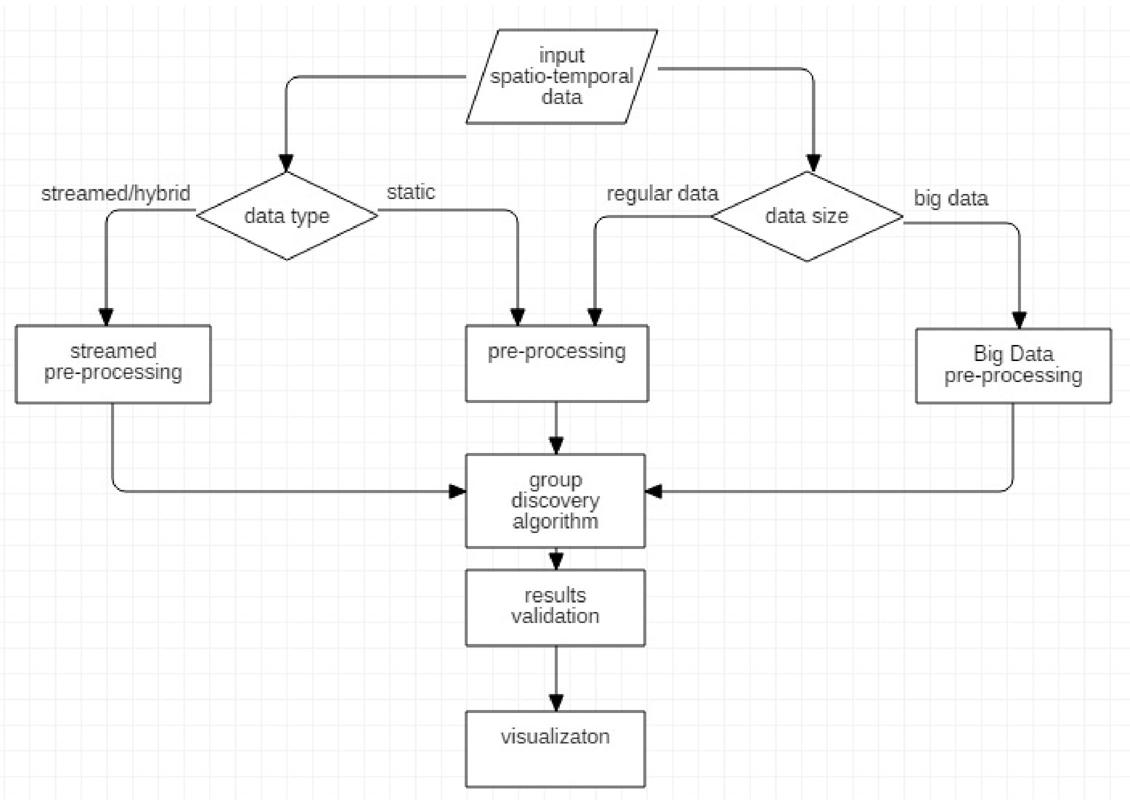
5 Research Methodology

In order to achieve the dissertation goals, an extensive literature review will be conducted. Since a significant majority of the published work in this and related topics do not include the source code, nor the original datasets, this work will be primarily based on secondary sources from the existing publications in the discussed subject.

The proposed dissertation will take a form of a survey/taxonomy of the existing group discovery models from spatio-temporal trajectories. In order to be able to perform a comprehensive classification of these models, a conceptual framework will be proposed that encapsulates all the reviewed features and processes leading to group discoveries. This will be based on a top-down approach by formulating a general concept and overview of the group pattern discovery from spatio-temporal trajectories and then refined based on the result of a bottom-up approach taken to define and classify the algorithms by decomposing their elements into parameters, influencing factors, etc.

A revision of the existing frameworks and approaches to data mining in spatial or spatio-temporal trajectories leads to a conclusion that these frameworks are either simplified and not including the details of the whole process which means they cannot be generalized to include other models, or they are not specifically designed for group pattern mining. Three most commonly used frameworks for group pattern discovery which are similar if not the same in their essence are: filter-refinement, partition and grouping and clustering and intersection. Filter and refinement are used for the classical algorithms such as flock and convoy and with certain amendments in more recent patterns such as traveling companion [Tang et al., 2013] but do not take into consideration the input data which may require for instance big data processing. Partition and grouping is a similar approach used in trajectory clustering [Lee et al., 2007] while clustering and intersection are mentioned in several publications on various approaches [Zheng, 2015, Gudmundsson and Horton, 2016, Jeung et al., 2010]. In *Spatiotemporal Data Mining: A Computational Perspective*, Shekhar et al. introduce a framework for the data mining from spatio-temporal trajectories [Shekhar et al., 2015], however this process does not include different input data such as static, streamed, or big data which change the way of pre-processing data in order to lead to group discovery. Another attempt to introduce a unified theoretical framework for data mining in general was made by Khan et al. [Khan et al., 2013] which included key elements such as data sets, cluster sets set of rules and visualization, however again did not differentiate between the data type and data size, nor allowed other methods than clustering (which are for instance a case of the original flock algorithm which introduced disk based group formation rather than density clustering known from later approaches).

A review conducted so far aimed at sketching an overview of the existing algorithms, trying to capture their commonalities as well as differences. The currently drafted framework based on the literature review of 61 papers is presented below:



While this is only a draft model which is subject to changes, it provides a basic structure for further analysis and eventually classify the discussed algorithms. In this model there are three different pre-processing stages which are dependent on the input data type and data size. The algorithms can be classified based on whether they scale up and are compatible with streaming data or not and whether they require a distributed or parallel programing on not. Next step is the pre-processing stage which includes all forms of clustering or filtering as per aforementioned frameworks and then again depending on the chosen pre-processing method the algorithms can be classified accordingly. The group discovery algorithm would then include all sort of parameters such as duration, density, group size, frequency, time-gap, etc. which eventually lead to analysis and validation of the results based on the selected parameters. Visualization is an additional UI layer which may or may not be eventually included in the final framework.

6 Research Schedule

The proposed dissertation will start from the week commencing on 01/06/2018 and will last until 16/08/2018. The below Gantt chart shows the split of tasks grouped by:

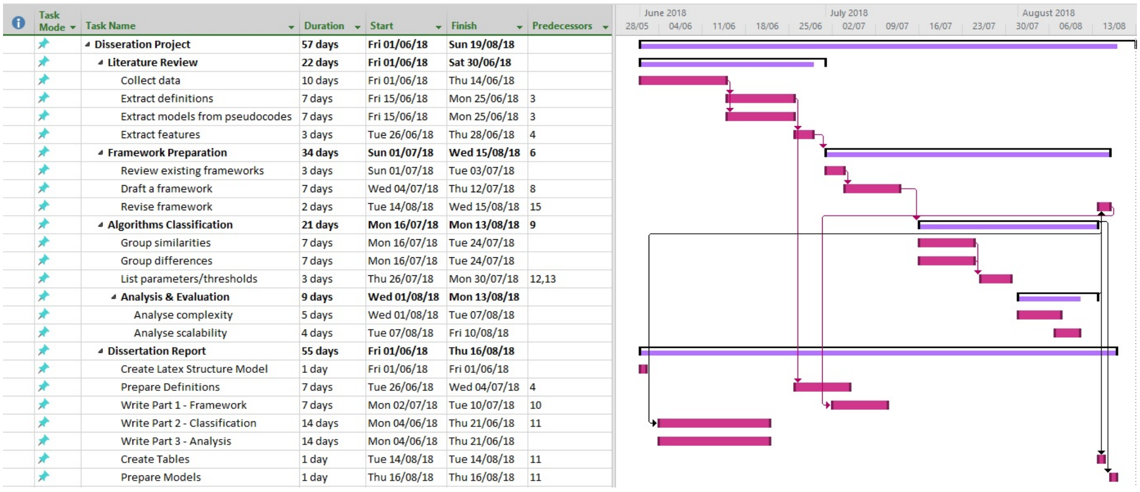
1. Literature Review
2. Framework Preparation
3. Algorithms Classification
4. Analysis and Evaluation
5. Dissertation Report

Each of the tasks is split into workable units.

Task Mode	Task Name	Duration	Start	Finish	Predecessors
	Disseration Project	**57 days**	**Fri 01/06/18**	**Sun 19/08/18**	
	Literature Review	**22 days**	**Fri 01/06/18**	**Sat 30/06/18**	
	Collect data	10 days	Fri 01/06/18	Thu 14/06/18	
	Extract definitions	7 days	Fri 15/06/18	Mon 25/06/18	3
	Extract models from pseudocodes	7 days	Fri 15/06/18	Mon 25/06/18	3
	Extract features	3 days	Tue 26/06/18	Thu 28/06/18	4
	Framework Preparation	**34 days**	**Sun 01/07/18**	**Wed 15/08/18**	**6**
	Review existing frameworks	3 days	Sun 01/07/18	Tue 03/07/18	
	Draft a framework	7 days	Wed 04/07/18	Thu 12/07/18	8
	Revise framework	2 days	Tue 14/08/18	Wed 15/08/18	15
	Algorithms Classification	**21 days**	**Mon 16/07/18**	**Mon 13/08/18**	**9**
	Group similarities	7 days	Mon 16/07/18	Tue 24/07/18	
	Group differences	7 days	Mon 16/07/18	Tue 24/07/18	
	List parameters/thresholds	3 days	Thu 26/07/18	Mon 30/07/18	12,13
	Analysis & Evaluation	**9 days**	**Wed 01/08/18**	**Mon 13/08/18**	
	Analyse complexity	5 days	Wed 01/08/18	Tue 07/08/18	
	Analyse scalability	4 days	Tue 07/08/18	Fri 10/08/18	

	Dissertation Report	**55 days**	**Fri 01/06/18**	**Thu 16/08/18**	
	Create Latex Structure Model	1 day	Fri 01/06/18	Fri 01/06/18	
	Prepare Definitions	7 days	Tue 26/06/18	Wed 04/07/18	4
	Write Part 1 - Framework	7 days	Mon 02/07/18	Tue 10/07/18	10
	Write Part 2 - Classification	14 days	Mon 04/06/18	Thu 21/06/18	11
	Write Part 3 - Analysis	14 days	Mon 04/06/18	Thu 21/06/18	
	Create Tables	1 day	Tue 14/08/18	Tue 14/08/18	11
	Prepare Models	1 day	Thu 16/08/18	Thu 16/08/18	11

The dependencies between each tasks are shown as per below:



Task Mode	Task Name	Duration	Start	Finish	Predecessors
	Disseration Project	**57 days**	**Fri 01/06/18**	**Sun 19/08/18**	
	Literature Review	**22 days**	**Fri 01/06/18**	**Sat 30/06/18**	
	Collect data	10 days	Fri 01/06/18	Thu 14/06/18	
	Extract definitions	7 days	Fri 15/06/18	Mon 25/06/18	3
	Extract models from pseudocodes	7 days	Fri 15/06/18	Mon 25/06/18	3
	Extract features	3 days	Tue 26/06/18	Thu 28/06/18	4
	Framework Preparation	**34 days**	**Sun 01/07/18**	**Wed 15/08/18**	6
	Review existing frameworks	3 days	Sun 01/07/18	Tue 03/07/18	
	Draft a framework	7 days	Wed 04/07/18	Thu 12/07/18	8
	Revise framework	2 days	Tue 14/08/18	Wed 15/08/18	15
	Algorithms Classification	**21 days**	**Mon 16/07/18**	**Mon 13/08/18**	9
	Group similarities	7 days	Mon 16/07/18	Tue 24/07/18	
	Group differences	7 days	Mon 16/07/18	Tue 24/07/18	
	List parameters/thresholds	3 days	Thu 26/07/18	Mon 30/07/18	12,13
	Analysis & Evaluation	**9 days**	**Wed 01/08/18**	**Mon 13/08/18**	
	Analyse complexity	5 days	Wed 01/08/18	Tue 07/08/18	
	Analyse scalability	4 days	Tue 07/08/18	Fri 10/08/18	
	Dissertation Report	**55 days**	**Fri 01/06/18**	**Thu 16/08/18**	
	Create Latex Structure Model	1 day	Fri 01/06/18	Fri 01/06/18	
	Prepare Definitions	7 days	Tue 26/06/18	Wed 04/07/18	4
	Write Part 1 - Framework	7 days	Mon 02/07/18	Tue 10/07/18	10
	Write Part 2 - Classification	14 days	Mon 04/06/18	Thu 21/06/18	11
	Write Part 3 - Analysis	14 days	Mon 04/06/18	Thu 21/06/18	
	Create Tables	1 day	Tue 14/08/18	Tue 14/08/18	11
	Prepare Models	1 day	Thu 16/08/18	Thu 16/08/18	11

References

[Atluri et al., 2017] Atluri, G., Karpatne, A., and Kumar, V. (2017). Spatio-Temporal Data Mining: A Survey of Problems and Methods. 1(1):1–37.

[Benkert et al., 2008] Benkert, M., Gudmundsson, J., Hübner, F., and Wolle, T. (2008). Reporting flock patterns. *Computational Geometry: Theory and Applications*, 41(3):111–125.

[Dodge et al., 2008] Dodge, S., Weibel, R., and Lautensch??tz, A. K. (2008). Towards a taxonomy of movement patterns. In *Information Visualization*, volume 7, pages 240–252.

[Esling and Agon, 2012] Esling, P. and Agon, C. (2012). Time-series data mining. *ACM Computing Surveys*, 45(1):1–34.

[Gudmundsson, 2004] Gudmundsson, J. (2004). Efficient detection of motion patterns in spatio-temporal data sets. *Proceedings of the 12th annual ACM international workshop on Geographic information systems - GIS '04*, pages 250–257.

[Gudmundsson and Horton, 2016] Gudmundsson, J. and Horton, M. (2016). Spatio-Temporal Analysis of Team Sports – A Survey. 50(2):1–34.

[Jeung et al., 2008] Jeung, H., Shen, H. T., and Zhou, X. (2008). Convoy queries in spatio-temporal databases. *Proceedings - International Conference on Data Engineering*, pages 1457–1459.

[Jeung et al., 2010] Jeung, H., Yiu, M. L., Zhou, X., Jensen, C. S., and Shen, H. T. (2010). Discovery of Convoys in Trajectory Databases.

[Keogh et al., 2004] Keogh, E., Lonardi, S., and Ratanamahatana, C. A. (2004). Towards parameter-free data mining. In *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, page 206, Seattle, WA.

[Khan et al., 2013] Khan, D. M., Mohamudally, N., and Babajee, K. R. (2013). A Unified Theoretical Framework for Data Mining. *Procedia Computer Science*, 17:104–113.

[Lee et al., 2007] Lee, J.-G., Han, J., and Whang, K.-Y. (2007). Trajectory clustering: a partition-and-group framework. *Proceedings of the ACM SIGMOD*, pages 593–604.

[Li et al., 2010] Li, Z., Ding, B., Han, J., and Kays, R. (2010). Swarm: Mining relaxed temporal moving object clusters. *VLDB International Conference on Very Large Data Bases*, 3(1-2):723–734.

[Naserian et al., 2018] Naserian, E., Wang, X., Dahal, K., Wang, Z., and Wang, Z. (2018). Personalized location prediction for group travellers from spatial–temporal trajectories. *Future Generation Computer Systems*, 83:278–292.

[Orakzai et al., 2016] Orakzai, F., Calders, T., and Pedersen, T. B. (2016). Distributed convoy pattern mining. *Proceedings - IEEE International Conference on Mobile Data Management*, 2016-July:122–131.

[Shekhar et al., 2015] Shekhar, S., Jiang, Z., Ali, R., Eftelioglu, E., Tang, X., Gunturi, V., and Zhou, X. (2015). Spatiotemporal Data Mining: A Computational Perspective. *ISPRS International Journal of Geo-Information*, 4(4):2306–2338.

[Tang et al., 2012] Tang, L. A., Zheng, Y., Yuan, J., Han, J., Leung, A., Hung, C. C., and Peng, W. C. (2012). On discovery of traveling companions from streaming trajectories. *Proceedings - International Conference on Data Engineering*, (3):186–197.

[Tang et al., 2013] Tang, L.-A., Zheng, Y., Yuan, J., Han, J., Leung, A., Peng, W.-C., and Porta, T. L. (2013). A framework of traveling companion discovery on trajectory data streams. *ACM Transactions on Intelligent Systems and Technology*, 5(1):1–34.

[Vatsavai et al., 2012] Vatsavai, R. R., Ganguly, A., Chandola, V., Stefanidis, A., Klasky, S., and Shekhar, S. (2012). Spatiotemporal data mining in the era of big spatial data. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data - BigSpatial '12*, pages 1–10, Redondo Beach, California. ACM.

[Zheng et al., 2013] Zheng, K. ., Zheng, Y. ., Yuan, J. ., Shang, S., Zheng, K., Zheng, Y., and Yuan, N. J. (2013). On Discovery of Gathering Patterns from Trajectories. 6544829:242–253.

[Zheng, 2015] Zheng, Y. U. (2015). Trajectory Data Mining : An Overview. *ACM Trans. Intell. Syst. Technol.*, 6(3):1–41.