

Configuration Manual

MSc Research Project
Data Analytics

James Nagle
Student ID: x20191014

School of Computing
National College of Ireland

Supervisor: Dr Mohammed Hasanuzzaman

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: James Nagle
Student ID: X20191014
Programme: MSc in Data Analytics **Year:** 2022
Module: MSc Research Project
Lecturer: Dr Mohammed Hasanuzzaman
Submission Due Date: 15th December 2022
Project Title: Identifying Key Players in the 1916 Rising using Centrality Measures and K-Shell Decomposition

Word Count: 422

Page Count: 2

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: 

Date: 13th December 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

James Nagle
Student ID: x20191014

1 Software and Libraries

The analysis of the 1916 Rising network was carried out using the python programming language. Python, IPython and a number of libraries needed for the project were accessed through Anaconda¹. The libraries used to complete this project were pandas for manipulating data frames and networkx for creating and analysing networks². The final network visualisations were created using Gephi³, an open-source software used for analysis.

2 Creating Extremists Edge List

An edge list containing 13,328 edges was created by reading the movement of extremists in Ireland reports and recording each observation of two individuals seen in contact with each other in a csv file. This file, extremists.csv, contains a column for the date of the observation and columns for the two individuals.

Four IPython notebooks were written to process this file and output networks and statistics. The first notebook, “1. Create Edge List.ipynb”, reorders each row in extremists.csv so that the names are in alphabetical order. If, for example, one edge consisted of the pair “Tom Clarke,P.H. Pearse” and another edge consisted of the pair “P.H. Pearse,Tom Clarke”, networkx would create a directed network. By alphabetically ordering each edge networkx instead returns an undirected network. This corrected file is called “extremists_alph.csv” and is used in the remaining three IPython notebooks.

3 Visualisations in Gephi

The four IPython notebooks can be opened and examined, or they can be run in any order. The extremists.csv and extremists_alph.csv files are provided in the project submission. If the second of these files is deleted then notebook 1 will need to be run first to recreate it.

Notebook 2 and 4 save the networks created within as graphml files which can be opened and examined in Gephi. These are provided in the graphml folder or can be generated by running notebook 2 and 4. On the Overview tab a layout can be selected and applied to the network. Modularity can be applied to the network through the Network Overview panel on the righthand window. The Appearance panel on the left can be used to recolour parts of the network based on community membership found by the modularity analysis. In the Preview

¹ <https://www.anaconda.com/products/distribution>

² <https://networkx.org/>

³ <https://gephi.org/>

tab further customisation can be applied, such as adding label names, changing their font and size, scaling the edge size and adding a background colour.

The visualisations used in the project report were created from these graphml files and saved as gephi files in the gephi folder, from where they can be reopened in Gephi.