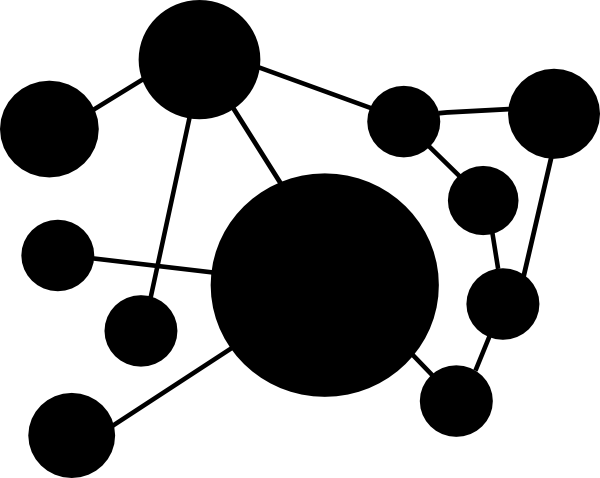


Modelling a Community:

Network Software and Parkrun Data





|  |  |
| --- | --- |
| Student Name (Student Number) | Eoin Carroll (16202781) |
| Module Code | UCD MIS40550 |
| Module Title | Network Software Modelling |
| Assessment Title | Network Software Modelling Assignment 2 |
| Module Co-ordinator | Dr James McDermott |
| Revision | 0 |
| Date Submitted | 23/04/2017 |

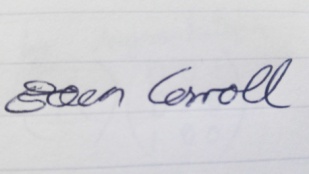
# Plagiarism

The unacknowledged inclusion of another person’s writings or ideas or works, in any formally presented work (including essays, examinations, projects, laboratory reports or presentations). The penalties associated with plagiarism designed to impose sanctions that reflect the seriousness of University’s commitment to academic integrity. Ensure that you have read the University’s Briefing for Students on Academic Integrity and Plagiarism and the UCD Plagiarism Statement, Plagiarism Policy and Procedures, (<http://www.ucd.ie/registrar/>).

# Declaration of Authorship

I declare that all material in this assessment is my own work except where there is clear acknowledgement and appropriate reference to the work of others.

# Signatures

Signed:  Date: 23/04/2017

# Abstract

Contents

[Introduction 4](#_Toc480321881)

[Graph Generation 4](#_Toc480321882)

[Web Scraping 4](#_Toc480321883)

[Constructing Graphs 4](#_Toc480321884)

[Combining Graphs 4](#_Toc480321885)

[Graph Properties 4](#_Toc480321886)

[Week 1 4](#_Toc480321887)

[Week 1&2 4](#_Toc480321888)

[Final Graph 4](#_Toc480321889)

[Information Flow 4](#_Toc480321890)

[Conclusion 4](#_Toc480321891)

[References 4](#_Toc480321892)

[Appendix 4](#_Toc480321893)

# Introduction

Parkrun . . . . a great way to get to know people and a large community has emerged. Also smaller communities at each location.

Information flow

Question – how will a message be passed

# Graph Generation

## Web Scraping

API offline during this project. Annoying but could be used in future, link.

Can apply for a research partnership but would have taken too long.

Manually copied latest 204 results set.

Could be run for others with API.

## Constructing Graphs

Read in excel, Parallel processing.

Same race, same name, same club, close position, close time. Within close time and position – name and club. Tested different parameters based mainly on domain knowledge, examined plots and adjusted.

Save as csv edgelist

Duplicate names not handles very well - just skip name. Skip unknowns

Run time was CPU intensive and took 3 hours. Most likely ways to speed up.

## Combining Graphs

Added all edges together and summed weights if the edges appeared more than once. Combine time was under 10mins, Ram intensive. Final file size

## Erdos-Renyi random graph model (Multiple Sizes)

# Graph Properties

Propose that anyone with a connection greater than 0.5 scaled knows each other.

Theoretical properties of Parkrun - all connected in each race. Possibility but unlikely that race would not have overlapping runners

Each race could be seen as a cluster

## Week 1

## Week 1 to 10

## Week 1 to 204

Full network – Marlay parkrun in it’s current state

## Graph Degree Plots

Degree Distribution (unscaled and log) for 5 graphs (First race will all be connected)

Degree Distribution with >0.5 for all 6 graphs

# Information Flow

Information passes based on scaled weight as a probability.

# Discussion

Note on run times

# Conclusion

# References

Clipart: http://www.clker.com/cliparts/G/G/F/Y/Y/U/network-md.png

Parkrun Logo: <http://www.renfrewshireleisure.com/media/187953/parkrun-logo.jpg>

Stackoverflow – I mostly google things

Run times of large graphs was significant

# Appendix

Include PC spec overclocked