Sporting contests through the lens of complex systems:

A data-driven approach

Joseph O'Brien

12-Mar-2021 | CIT Research Methods Seminar

UNIVERSITY of LIMERICK

OLLSCOIL LUIMNIGH

✉ joseph.obrien@ul.ie

@obrienj_

Fondúireacht Eolaíochta Éireann
Dá bhfuil romhainn

Science Foundation Ireland
For what's next

About me

- PhD student at the University of Limerick
- Generally interested in understanding complex systems via mathematical modelling + data science
- Social media dynamics, science of science, financial systems, sporting applications
- **webpage:** joeyobrien.ie | **twitter:** @obrienj_ | **github:** obrienjoey

About me

- PhD student at the University of Limerick
- Generally interested in understanding complex systems via mathematical modelling + data science
- Social media dynamics, science of science, financial systems, sporting applications
- **webpage:** joeyobrien.ie | **twitter:** @obrienj_ | **github:** obrienjoey

About me

- PhD student at the University of Limerick
- Generally interested in understanding complex systems via mathematical modelling + data science
- Social media dynamics, science of science, financial systems, sporting applications
- **webpage:** joeyobrien.ie | **twitter:** @obrienj_ | **github:** obrienjoey

About me

- PhD student at the University of Limerick
- Generally interested in understanding complex systems via mathematical modelling + data science
- Social media dynamics, science of science, financial systems, sporting applications
- **webpage:** joeyobrien.ie | **twitter:** @obrienj_ | **github:** obrienjoey

Outline of Today's Talk

- Publicly available data + webscraping
- Skill + **Fantasy Premier League** managers
- **Snooker** and its ranking problem
- **Conclusions**

Outline of Today's Talk

- Publicly available data + webscraping
- Skill + **Fantasy Premier League** managers
- **Snooker** and its ranking problem
- **Conclusions**

Outline of Today's Talk

- Publicly available data + webscraping
- Skill + **Fantasy Premier League** managers
- **Snooker** and its ranking problem
- **Conclusions**

Outline of Today's Talk

- Publicly available data + webscraping
- Skill + **Fantasy Premier League** managers
- **Snooker** and its ranking problem
- **Conclusions**

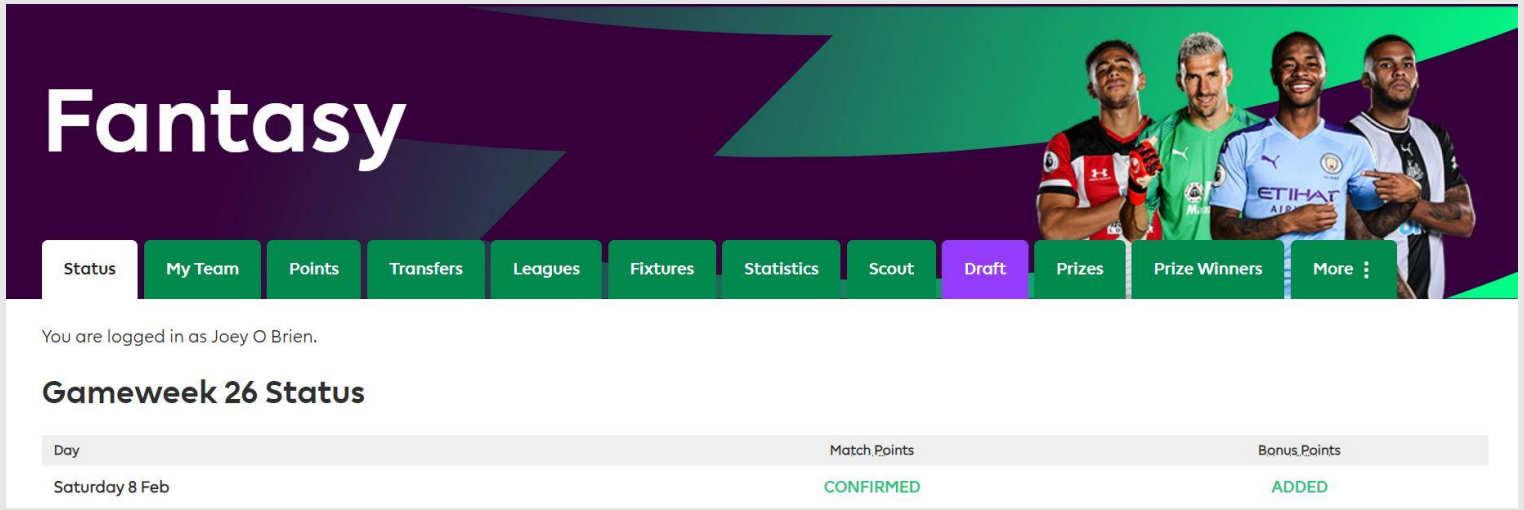
Outline of Today's Talk

- Publicly available data + webscraping
- Skill + **Fantasy Premier League** managers
- **Snooker** and its ranking problem
- **Conclusions**

What is Webscraping All About?

- The web is essentially a repository for data.

- In order to store and display that data the websites have to be coded - generally in **HTML** or **Javascript**.



What is Webscraping All About?

- The web is essentially a repository for data.

- In order to store and display that data the websites have to be coded - generally in **HTML** or **Javascript**.

<!doctype html><html lang="en"><head><meta charset="utf-8"/><meta http-equiv="x-ua-compatible" content="ie=edge"/><meta name="viewport" content="width=device-width,initial-scale=1"/><title>Fantasy Premier League, Official Fantasy Football Game of the Premier League</title><meta name="description" content="Official Fantasy Premier League 2019/20. Free to play fantasy football game, set up your fantasy football team at the Official Premier League site." data-react-helmet="true"/><link rel="apple-touch-icon" sizes="180x180" href="/img/favicons/apple-touch-icon.png"/><link rel="icon" type="image/png" href="/img/favicons/favicon-32x32.png" sizes="32x32"/><link rel="icon" type="image/png" href="/img/favicons/favicon-16x16.png" sizes="16x16"/><link rel="manifest" href="/img/favicons/manifest.json"/><link rel="mask-icon" href="/img/favicons/safari-pinned-tab.svg" color="#5bbad5"/><link rel="shortcut icon" href="/img/favicons/favicon.ico"/><meta name="msapplication-config" content="/favicons/browserconfig.xml"/><meta name="theme-color" content="#ffffff"/><meta property="og:url" content="https://fantasy.premierleague.com/"/><meta property="og:type" content="website"/><meta property="og:title" content="Fantasy Premier League, Official Fantasy Football Game of the Premier League"/><meta property="og:description" content="Official Fantasy Premier League 2019/20. Free to play fantasy football game, set up your fantasy football team at the Official Premier League site."/><meta property="og:image" content="/img/share/facebook-share.png"/><meta name="twitter:card" content="summary"/><meta name="twitter:site" content="@OfficialFPL"/><meta name="apple-itunes-app" content="app-id=1138895159"/><link rel="stylesheet" href="//www.premierleague.com/resources/prod/a83e4e7-1591/ism/css/ism.css"/><script async src="https://platform.twitter.com/widgets.js"></script><script type="text/javascript">var googletag=googletag||{};googletag.cmd=googletag.cmd||[],function(){var t=document.createElement("script");t.async=!0,t.type="text/javascript";var e="https:"==document.location.protocol;t.src=(e?"https:":"http:")+"//www.googletagservices.com/tag/js/gpt.js";var o=document.getElementsByTagName("script")[0];o.parentNode.insertBefore(t,o)}()</script><script>!function(e,t,n,c,o,a,f){e.fbq||(o=e.fbq=function(){o.callMethod?o.callMethod.apply(o,arguments):o.queue.push(arguments)},e._fbq||(e._fbq=o),(o.push=o).loaded=!0,o.version="2.0",o.queue=[],(a=t.createElement(n)).async=!0,a.src="https://connect.facebook.net/en_US/fbevents.js",(f=t.getElementsByTagName(n)

What is Webscraping All About?

- The web is essentially a repository for data.
- In order to store and display that data the websites have to be coded - generally in **HTML** or **Javascript**.
- Once we become comfortable with the syntax of these languages any data on the web is obtainable!

Why Webscrape?

- Many important publications use novel datasets to determine previously unconsidered phenomena
  - Think Stanley Milgram's research into the *six degrees of separation*
  - One of the many studies into citation dynamics

- Some fields suffer from 'Zachary Karate Club Syndrome'

Why Webscrape?

- Many important publications use novel datasets to determine previously unconsidered phenomena
  - Think Stanley Milgram's research into the *six degrees of separation*
  - One of the many studies into citation dynamics

- Some fields suffer from 'Zachary Karate Club Syndrome'

Why Webscrape?

- Many important publications use novel datasets to determine previously unconsidered phenomena
  - Think Stanley Milgram's research into the *six degrees of separation*
  - One of the many studies into citation dynamics

- Some fields suffer from 'Zachary Karate Club Syndrome'

Why Webscrape?

- Many important publications use novel datasets to determine previously unconsidered phenomena
  - Think Stanley Milgram's research into the *six degrees of separation*
  - One of the many studies into citation dynamics

- Some fields suffer from 'Zachary Karate Club Syndrome'

Zachary Karate Club Club [edit]

Zachary Karate Club Club is a honorific group[5] that awards membership in the group, along with a traveling trophy, to a scientist who is the first to use Zachary's Karate Club as an example at a conference on networks. The first scientist to be awarded was Cristopher Moore[6] in 2013, at a conference at the Santa Fe Institute.

Why Webscrape?

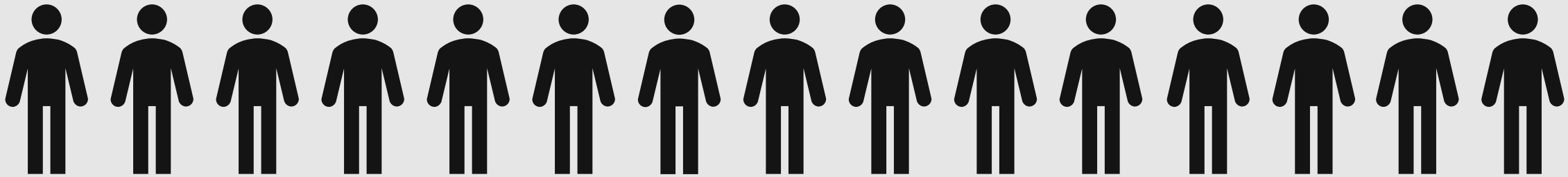
- Many important publications use novel datasets to determine previously unconsidered phenomena
  - Think Stanley Milgram's research into the *six degrees of separation*
  - One of the many studies into citation dynamics
- Some fields suffer from 'Zachary Karate Club Syndrome'
- A chance to do research on something that really interests you

1. Fantasy Premier League and the question of luck or skill

@obrienj_

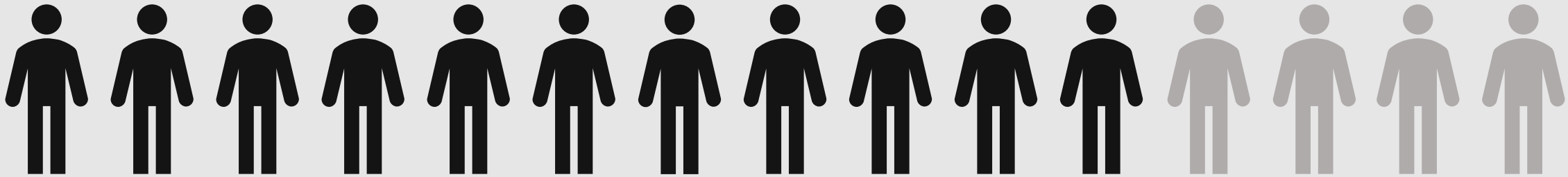
What is Fantasy Premier League?

- Online game where users are provided with a virtual £100 million budget to assemble a squad of 15 real-life Premier League footballers.
- Each week choosing 11 of them to ‘start’ for the user’s team, who then receive points based upon their statistical performance in the real game.
- Also choose one **captain** who receives double points. The challenge of the game is to **determine an optimal way of spending your budget as to maximize your points.**



What is Fantasy Premier League?

- Online game where users are provided with a virtual £100 million budget to assemble a squad of 15 real-life Premier League footballers.
- Each week choosing 11 of them to ‘start’ for the user’s team, who then receive points based upon their statistical performance in the real game.
- Also choose one **captain** who receives double points. The challenge of the game is to **determine an optimal way of spending your budget as to maximize your points.**



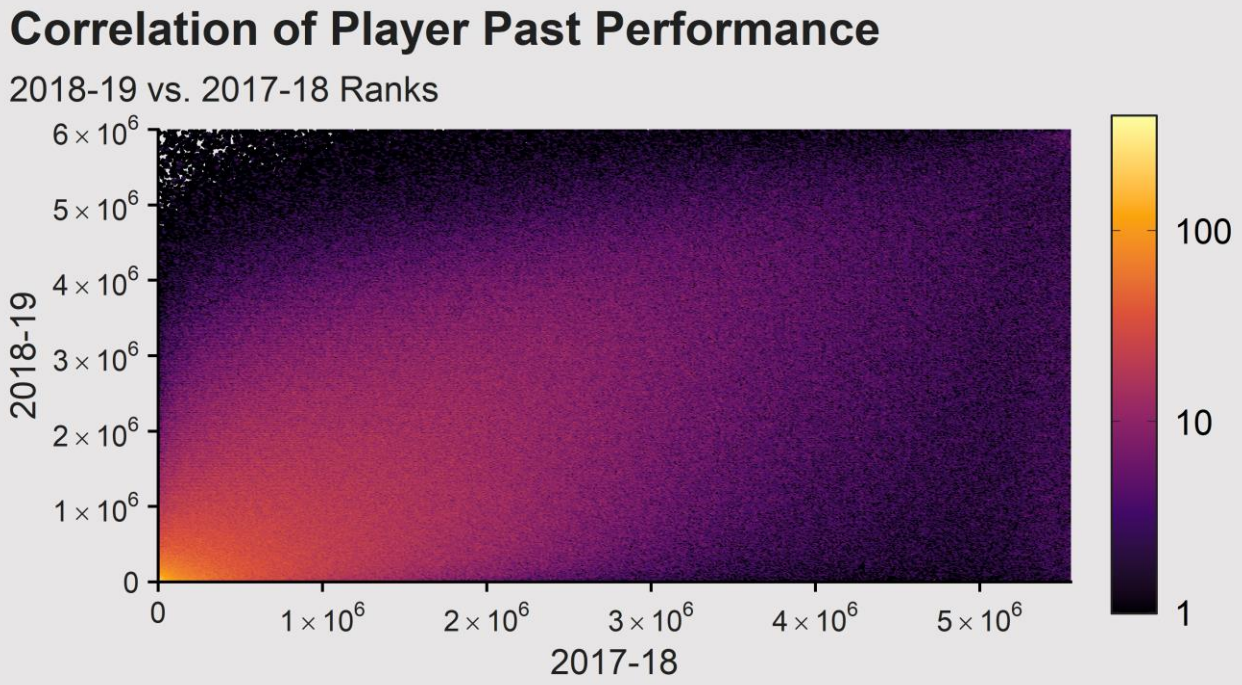
What is Fantasy Premier League?

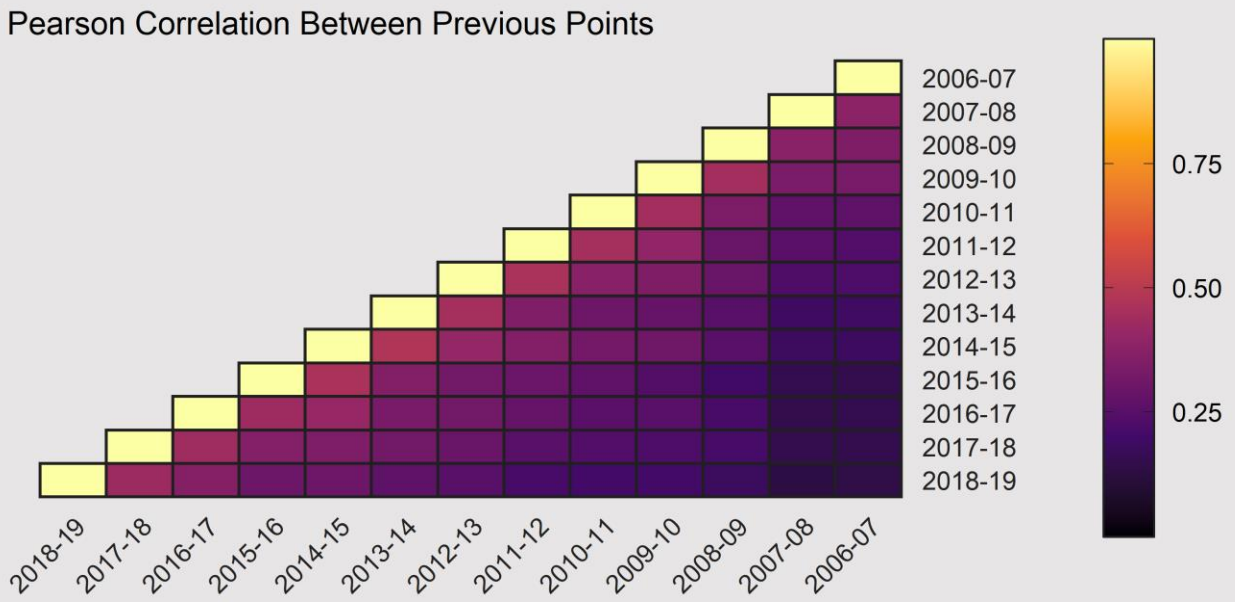
- Online game where users are provided with a virtual £100 million budget to assemble a squad of 15 real-life Premier League footballers.
- Each week choosing 11 of them to ‘start’ for the user’s team, who then receive points based upon their statistical performance in the real game.
- Also choose one **captain** who receives double points. The challenge of the game is to **determine an optimal way of spending your budget as to maximize your points.**

What is Fantasy Premier League?

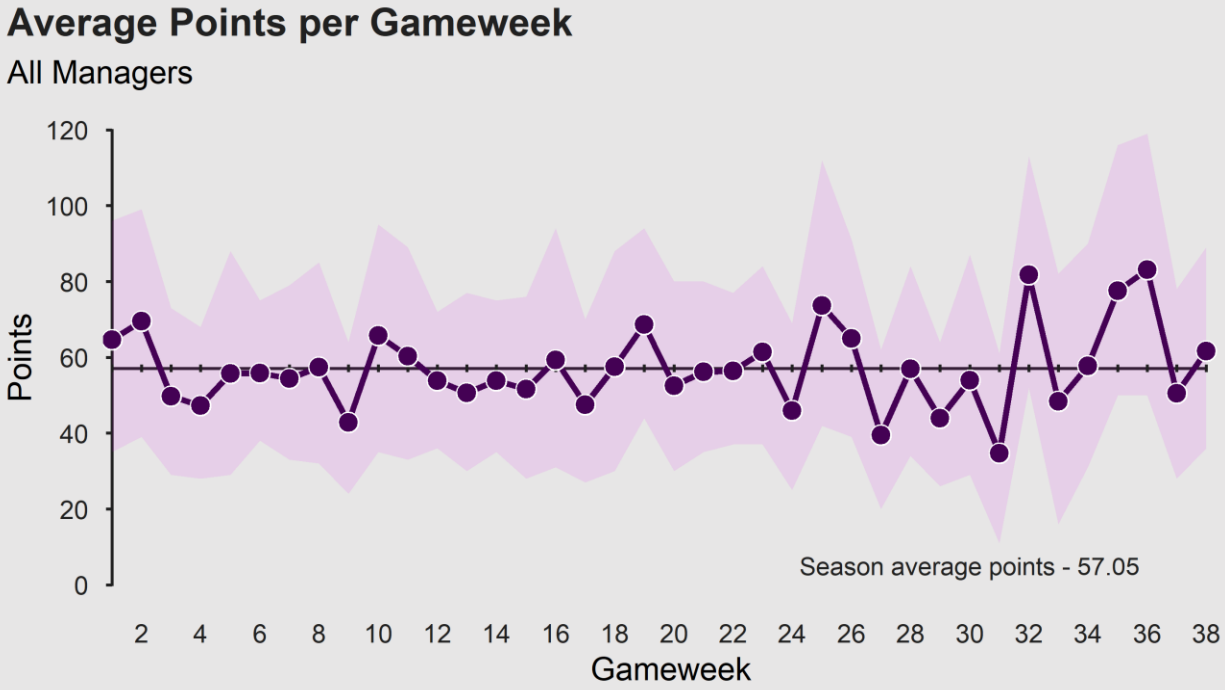
- Online game where users are provided with a virtual £100 million budget to assemble a squad of 15 real-life Premier League footballers.
- Each week choosing 11 of them to ‘start’ for the user’s team, who then receive points based upon their statistical performance in the real game.
- Also choose one **captain** who receives double points. The challenge of the game is to **determine an optimal way of spending your budget as to maximize your points.**

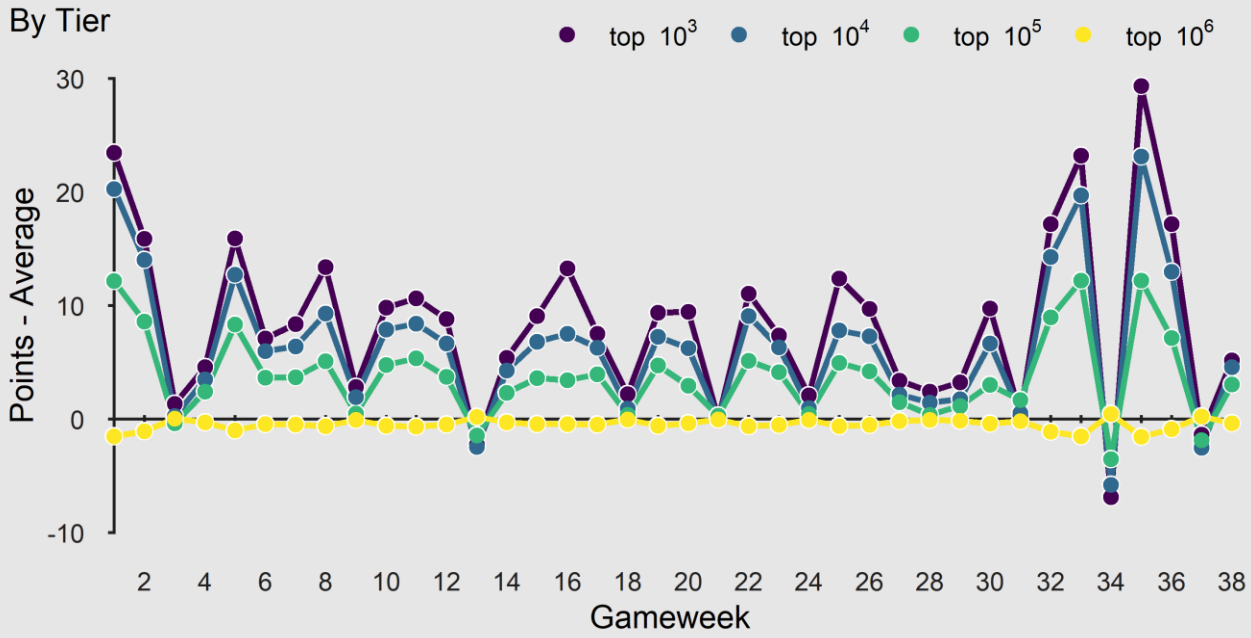
Consistency Between Performances?



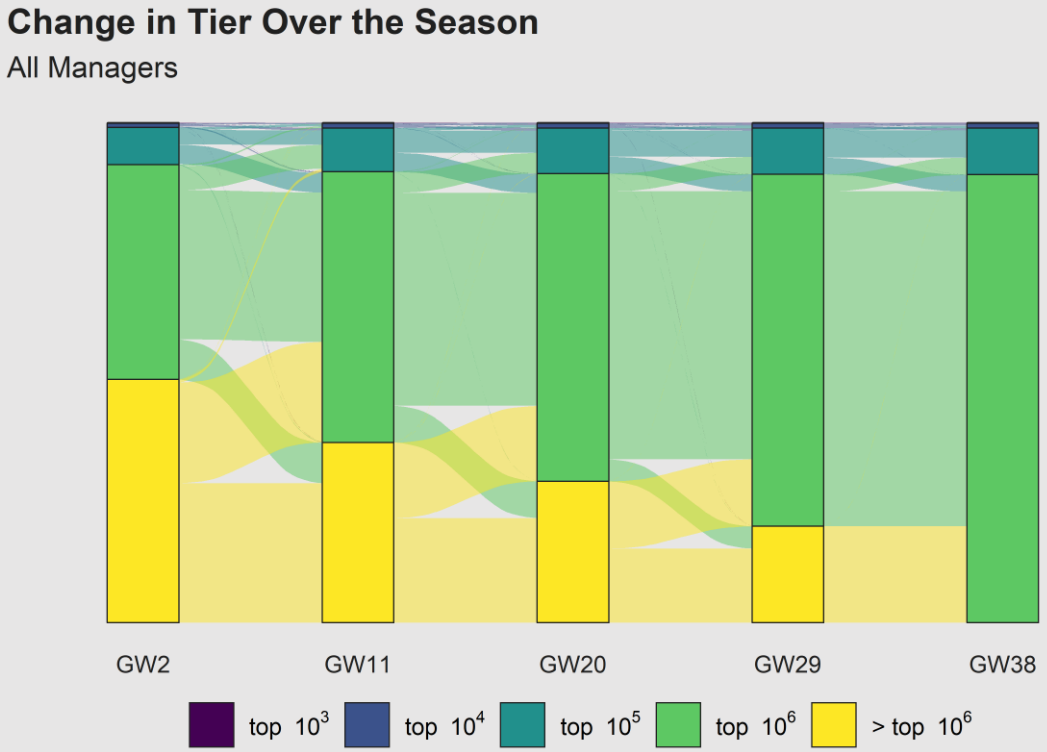


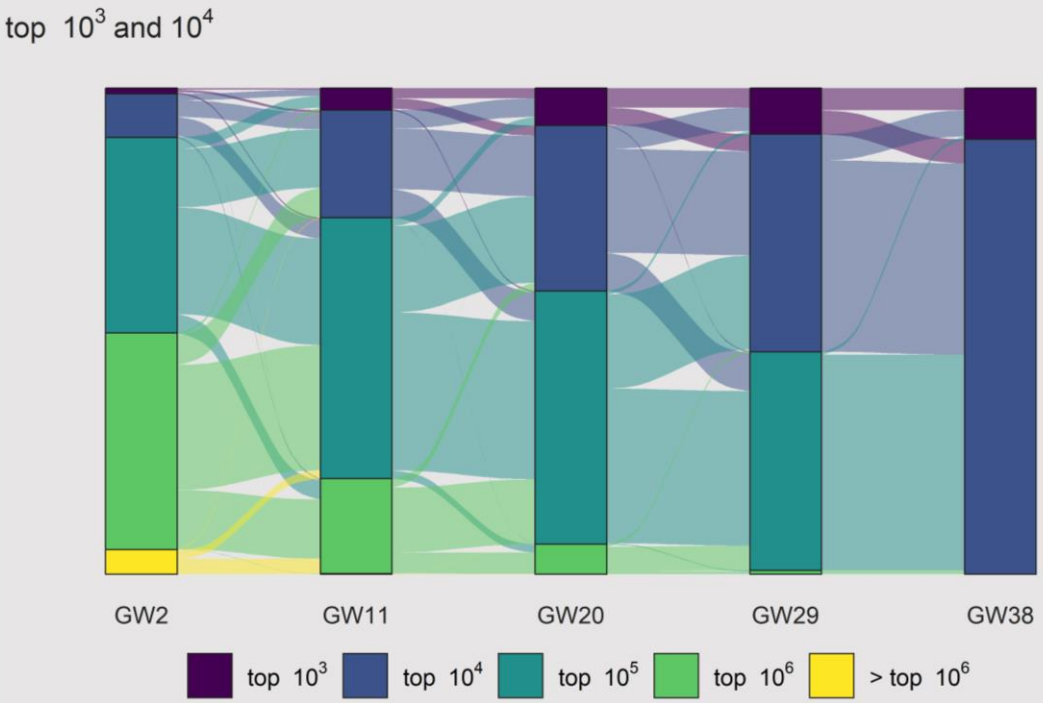
Tiers among managers



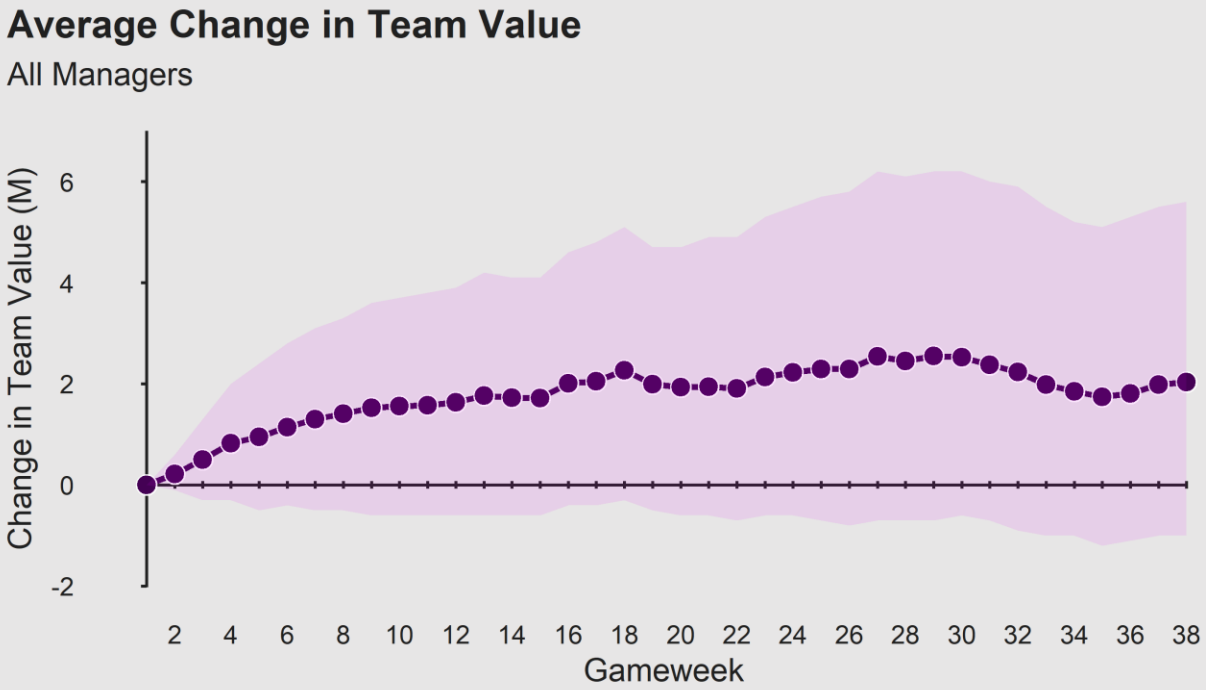


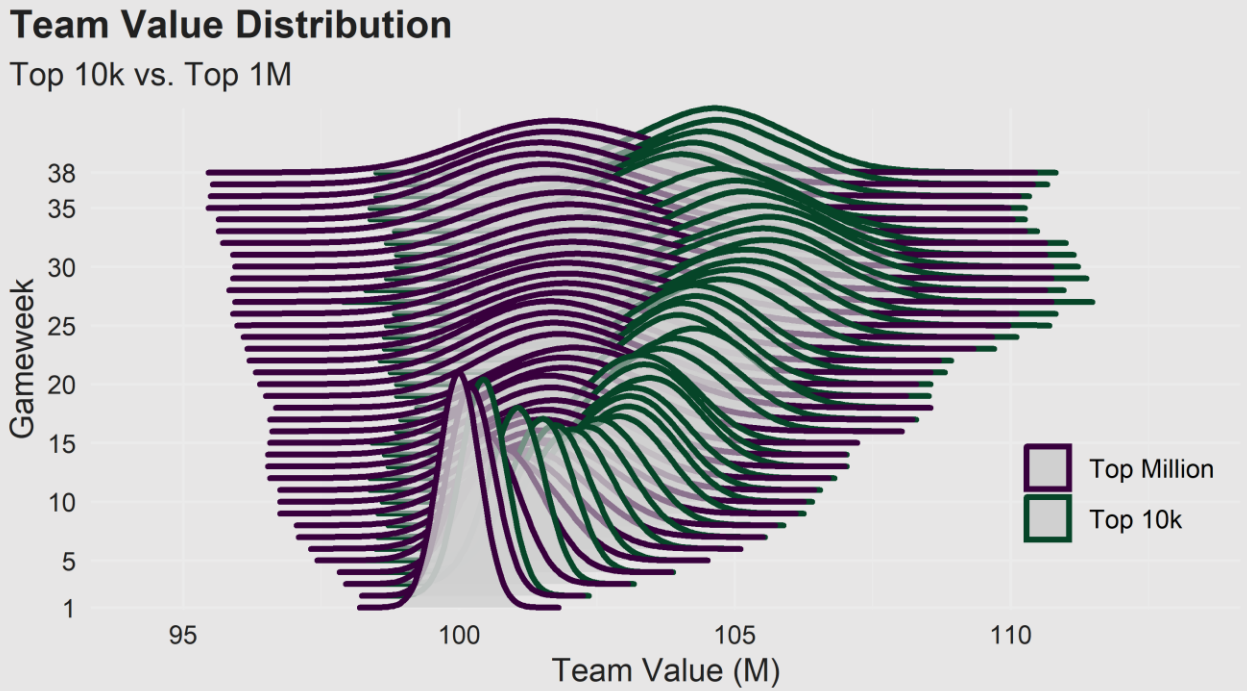
Tiers among managers



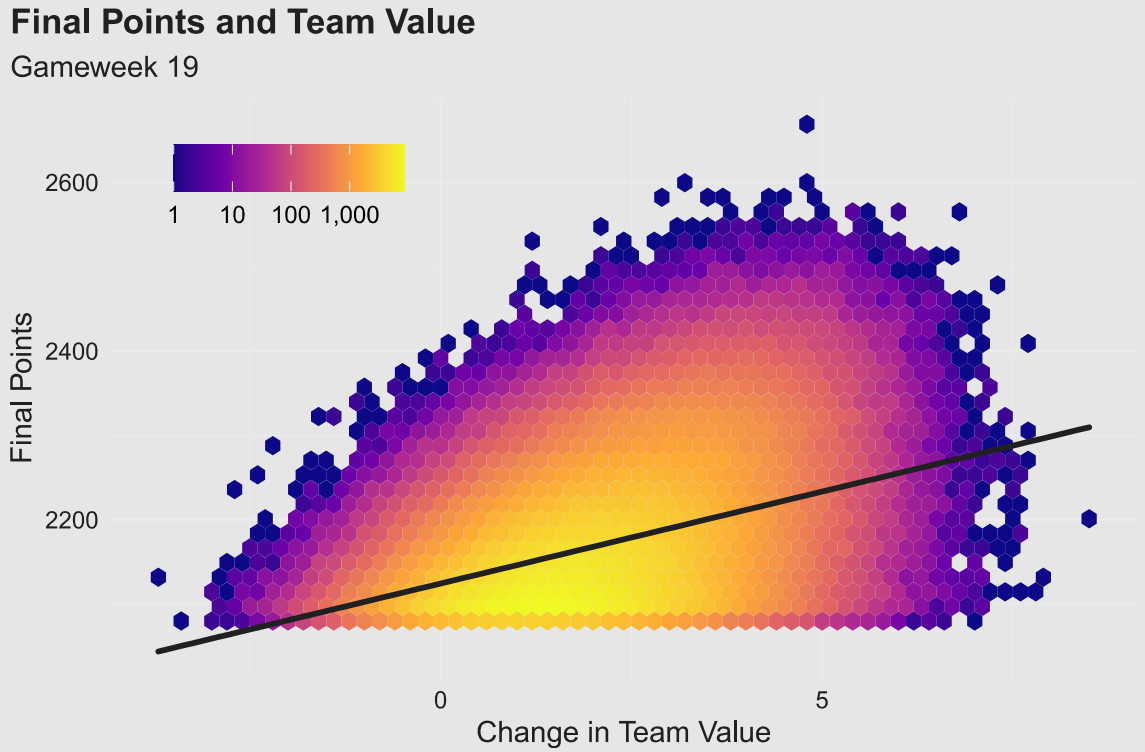


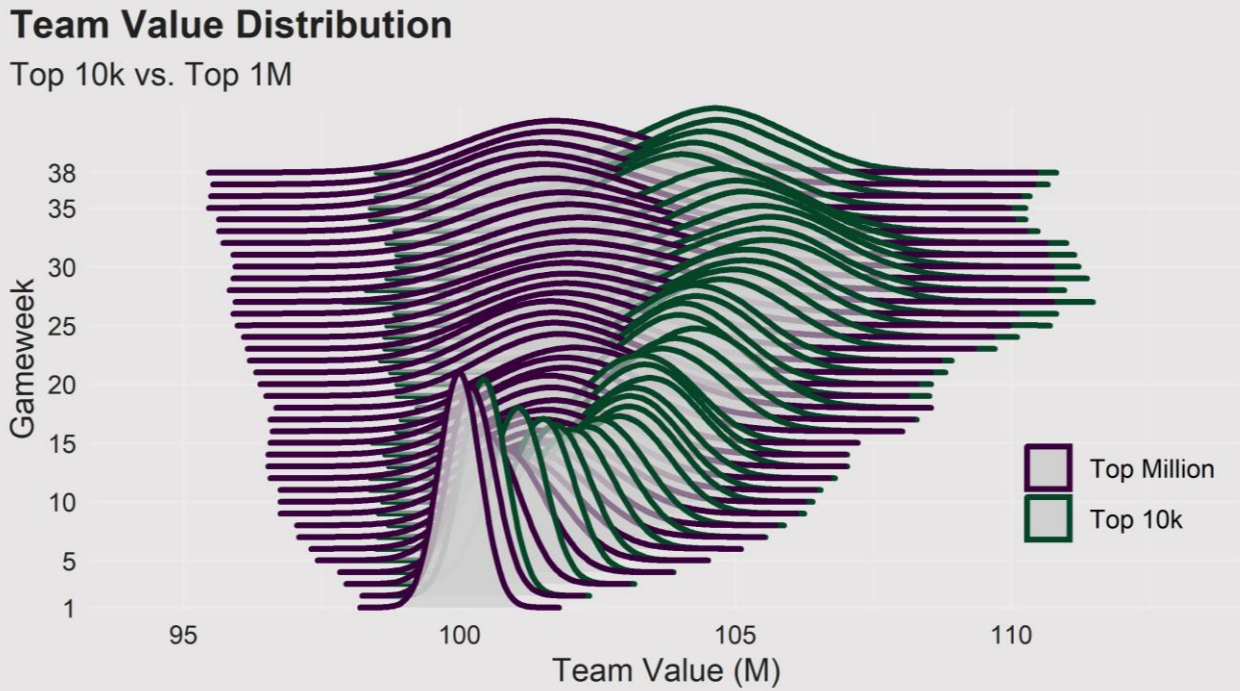
Difference in skill - money



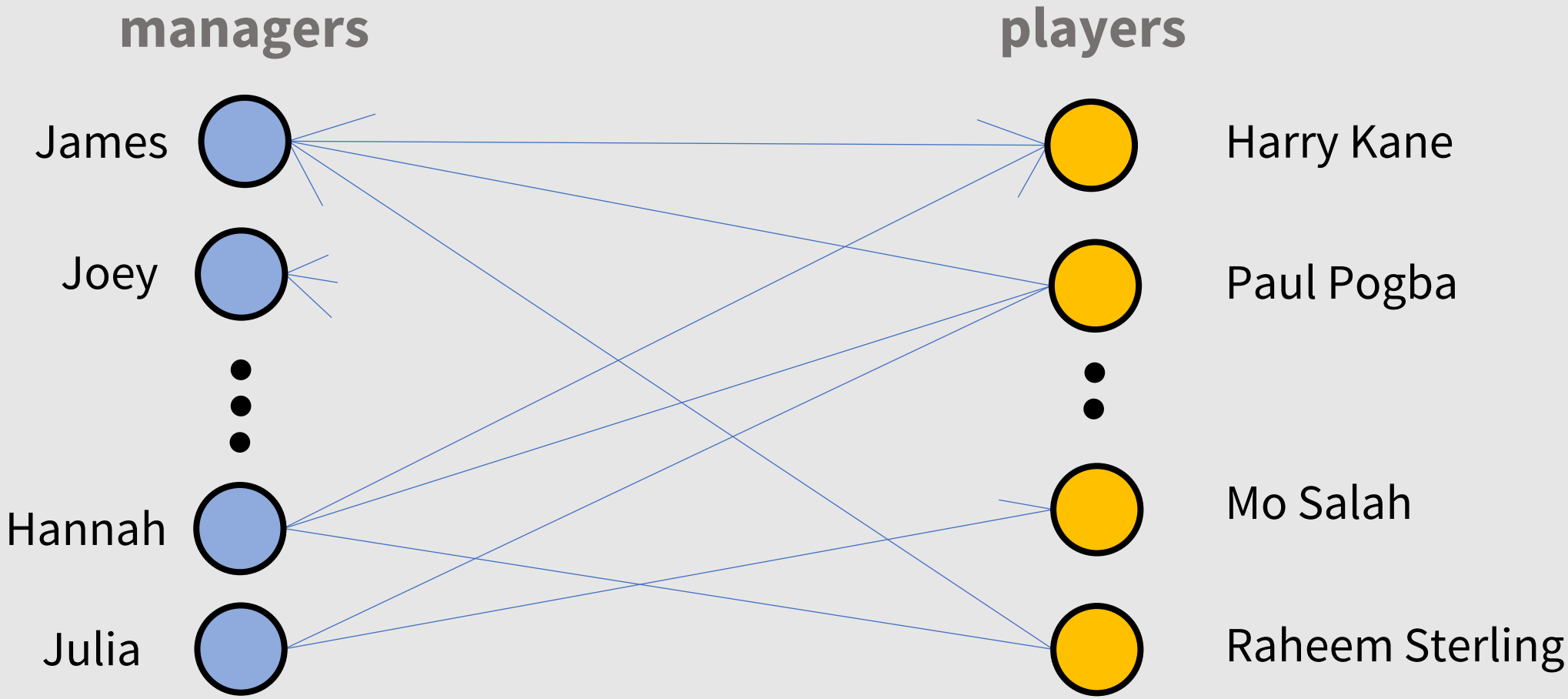


Difference in skill - money

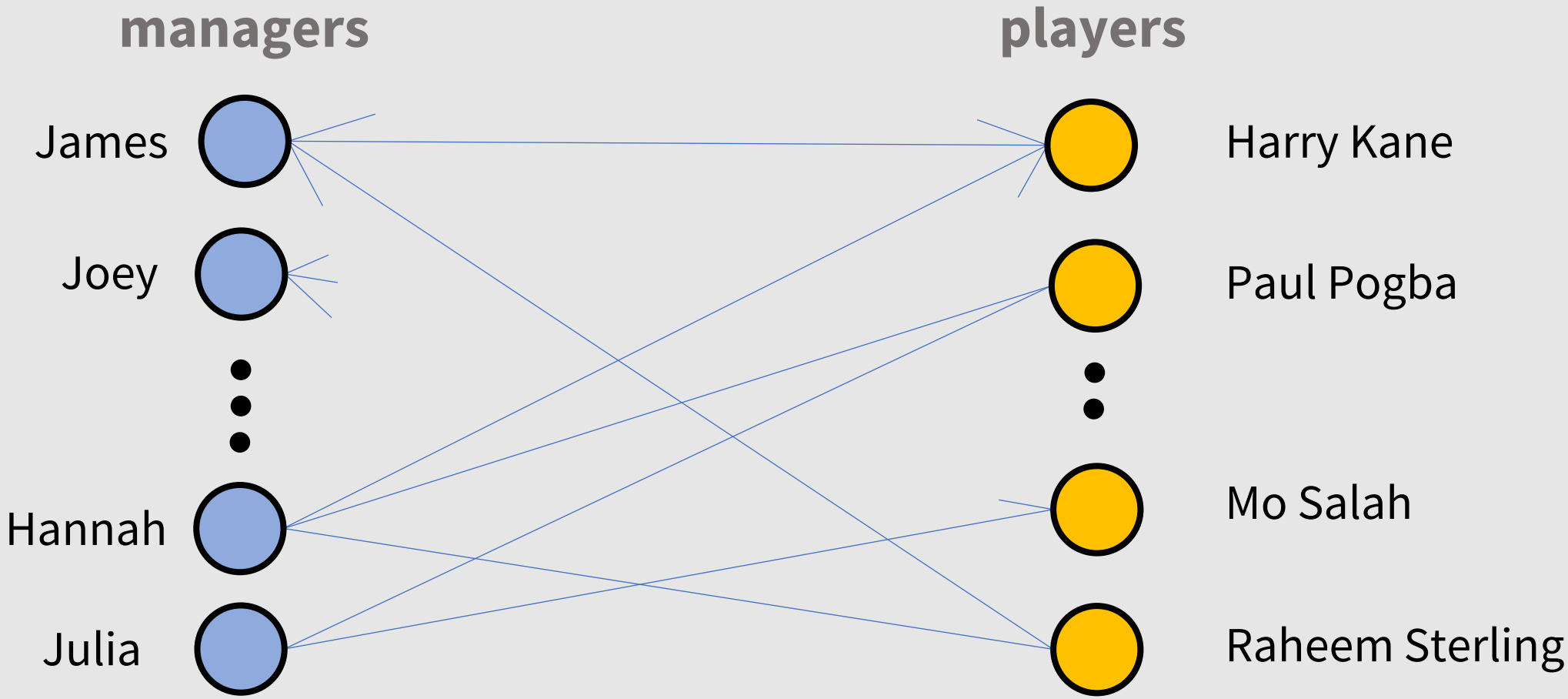




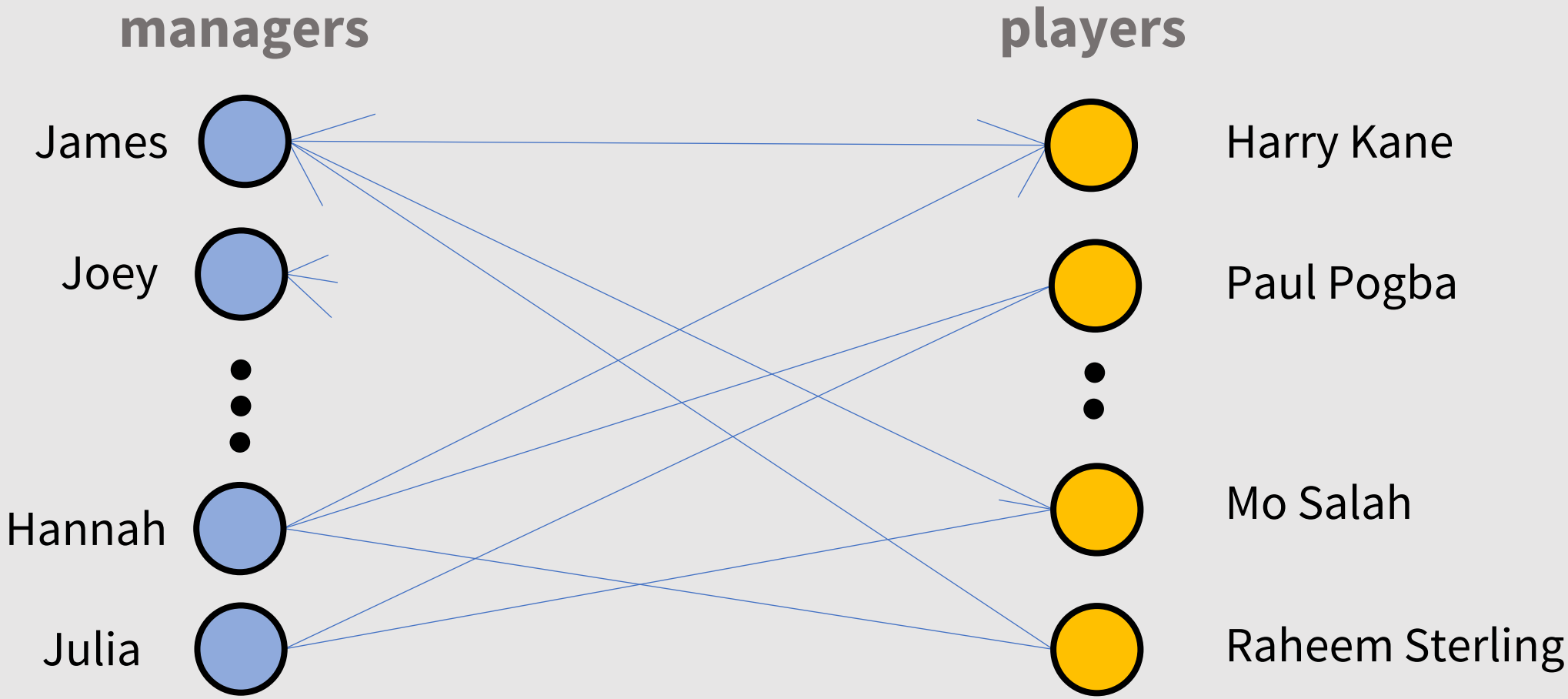
Network analysis



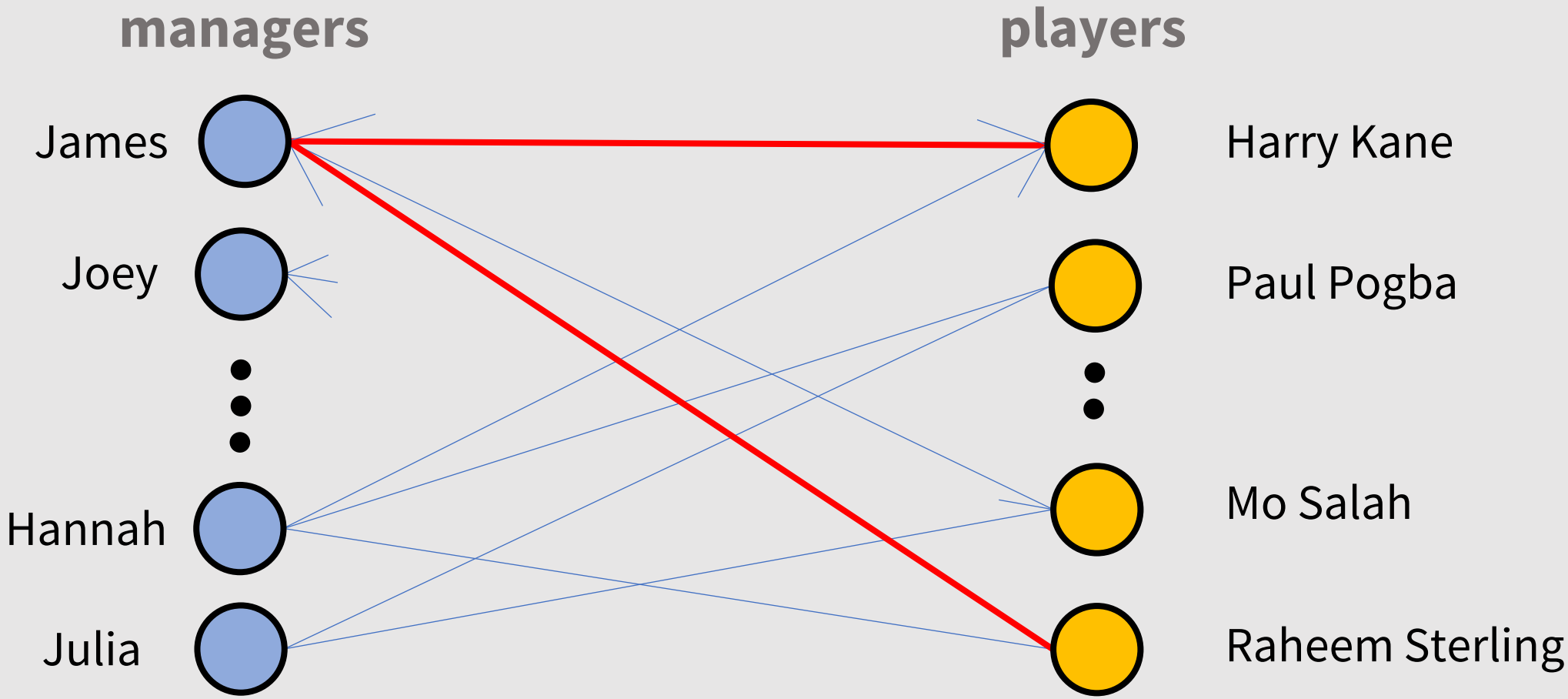
Network analysis



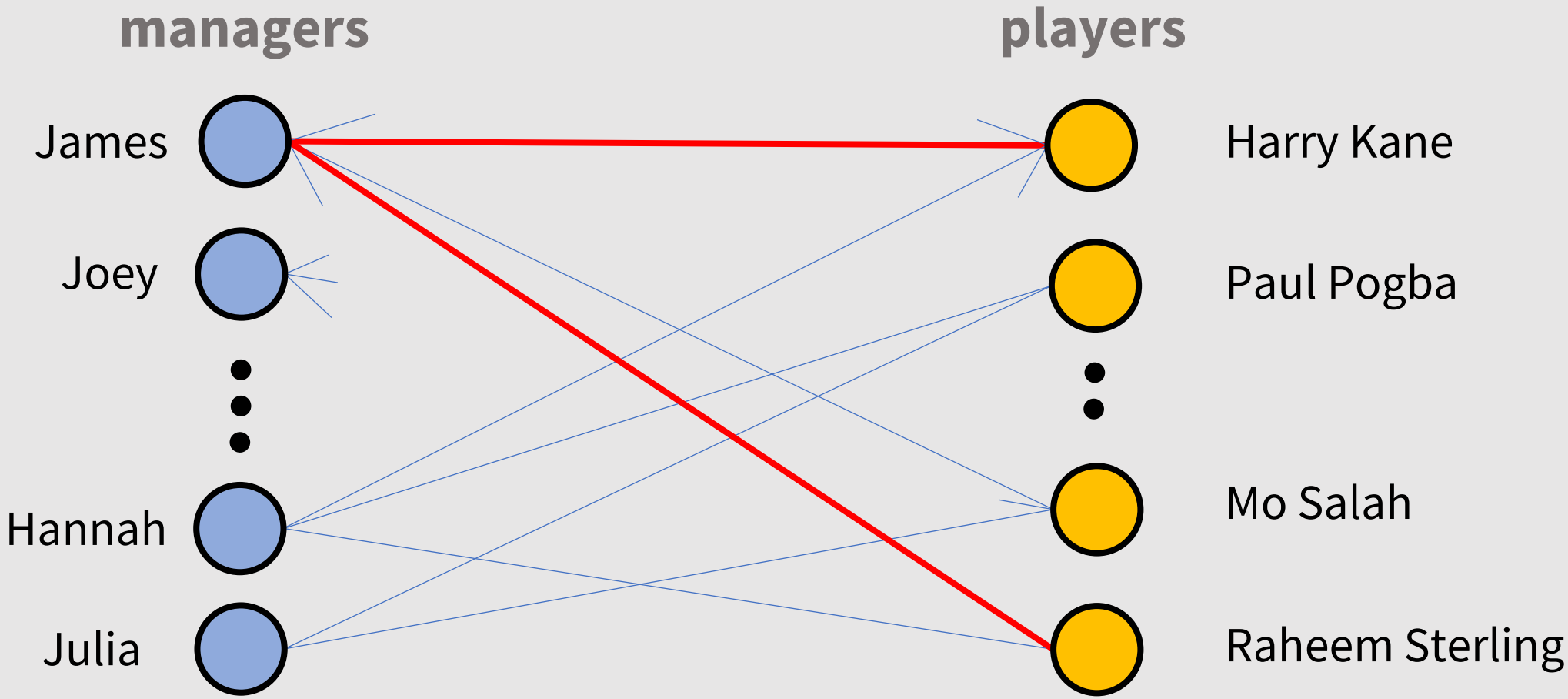
Network analysis



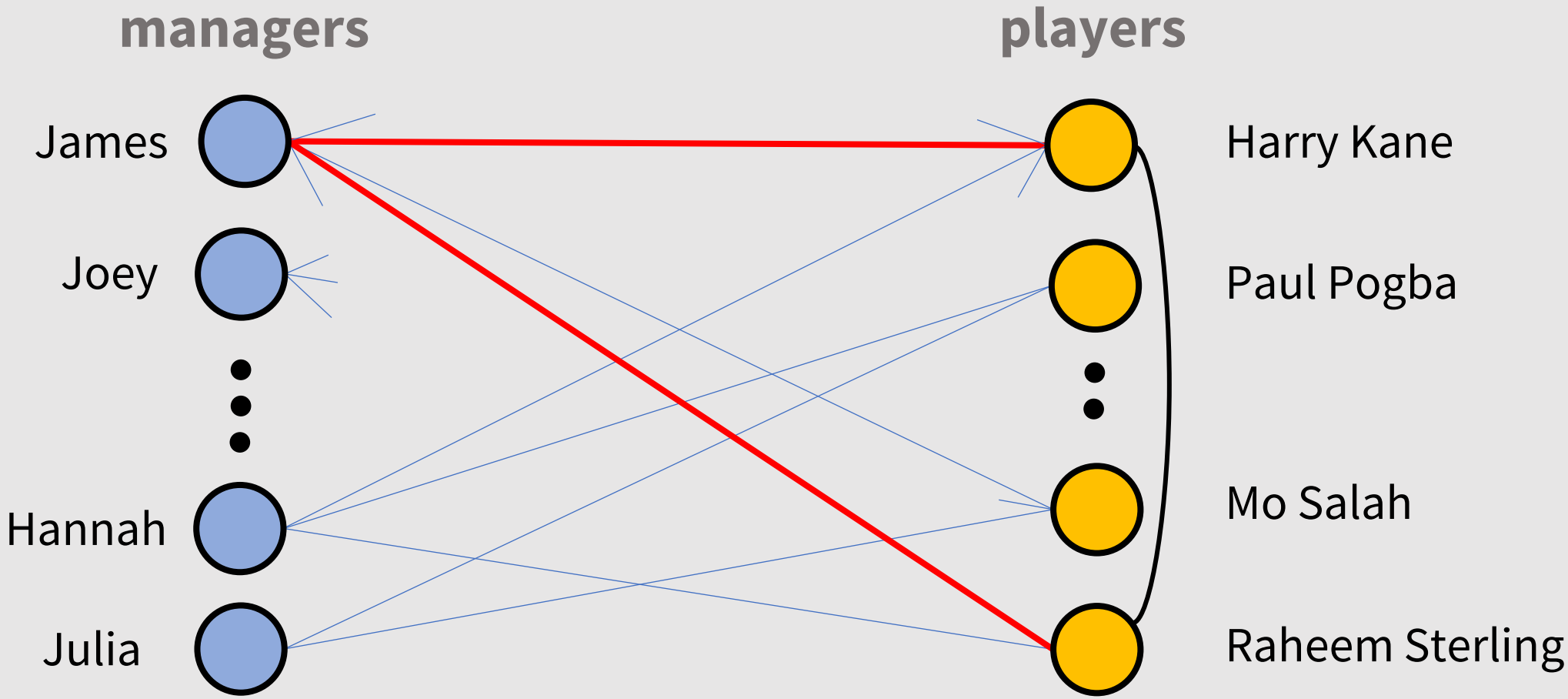
Network analysis



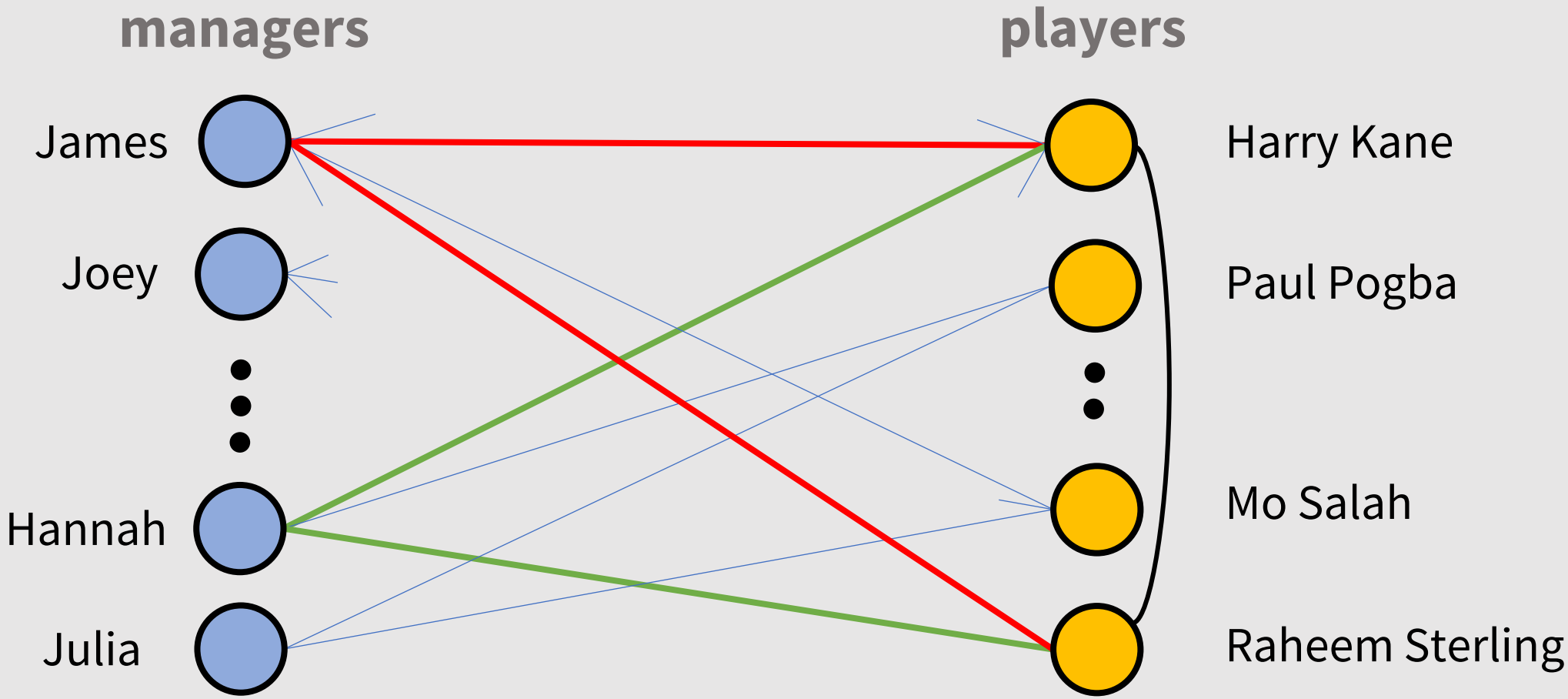
Network analysis



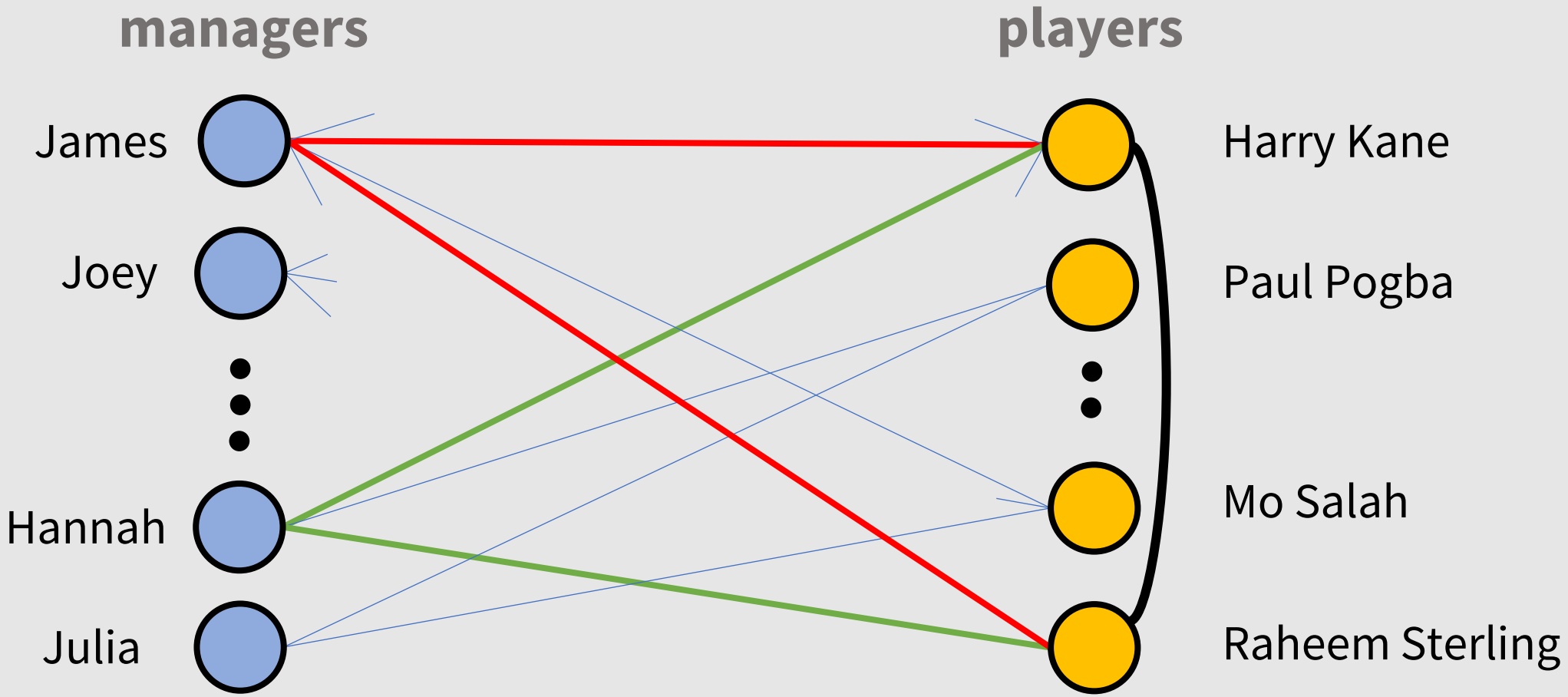
Network analysis



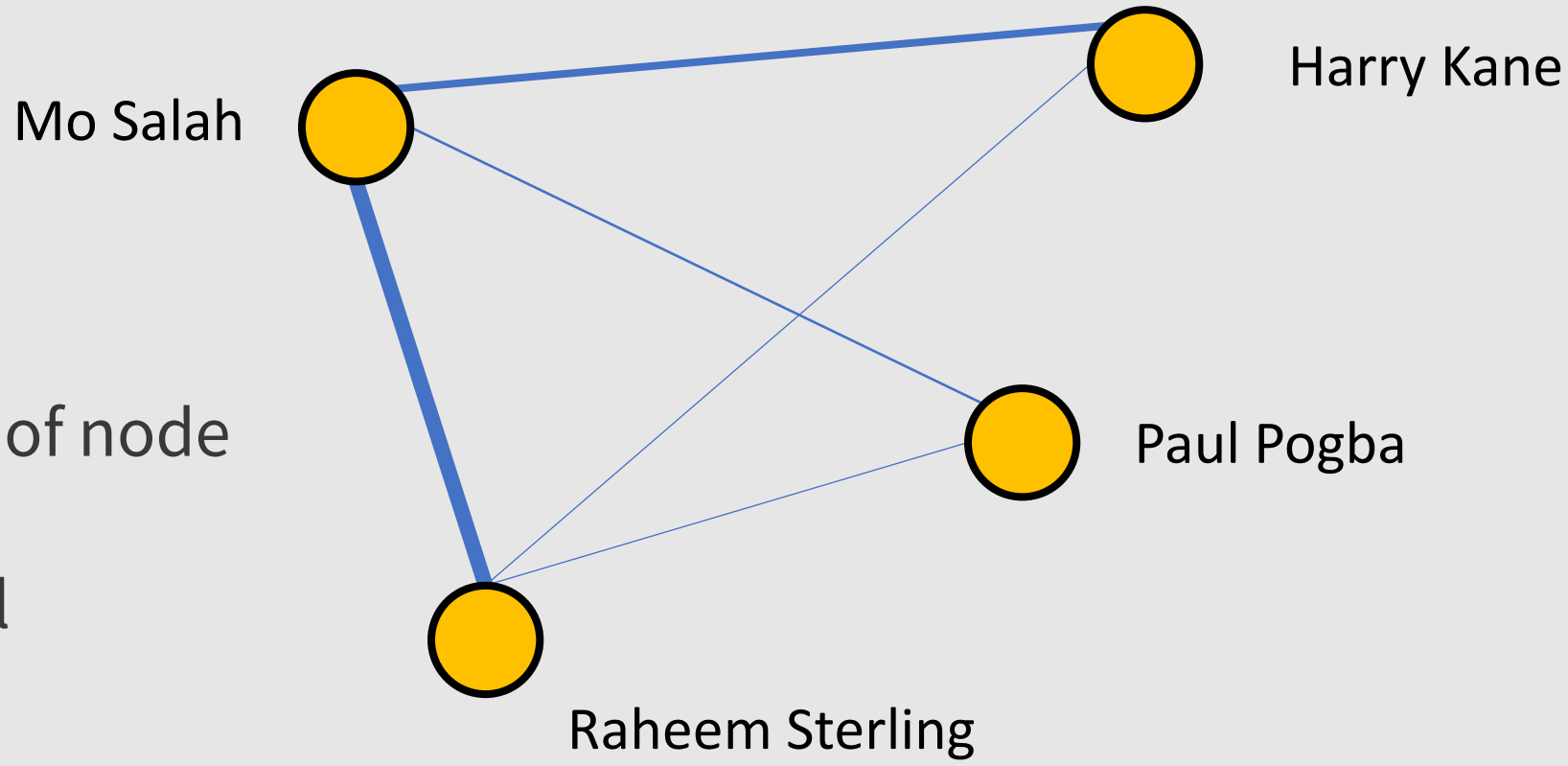
Network analysis



Network analysis



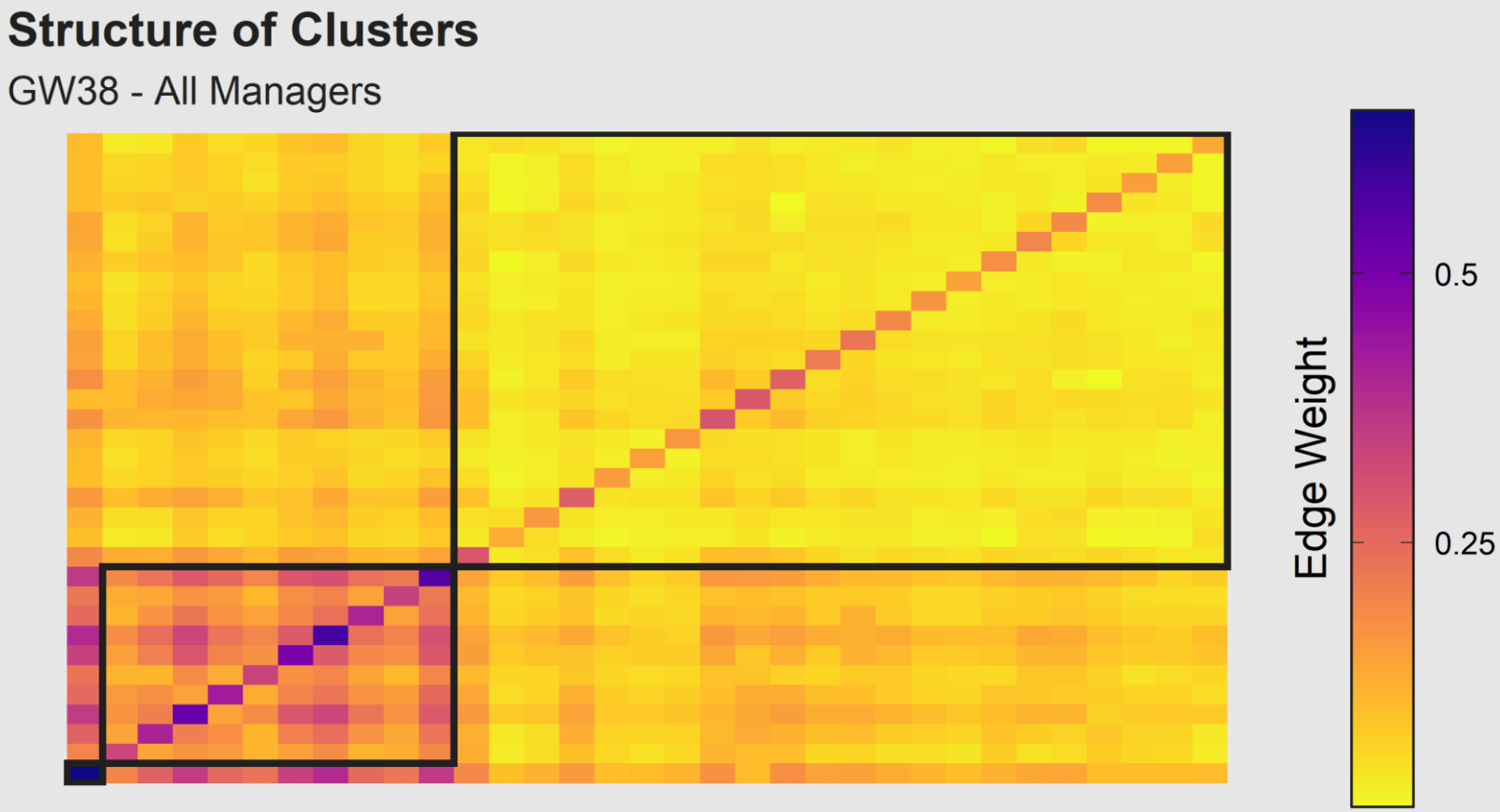
Network analysis



- One type of node
- Weighted
- Temporal

Template team

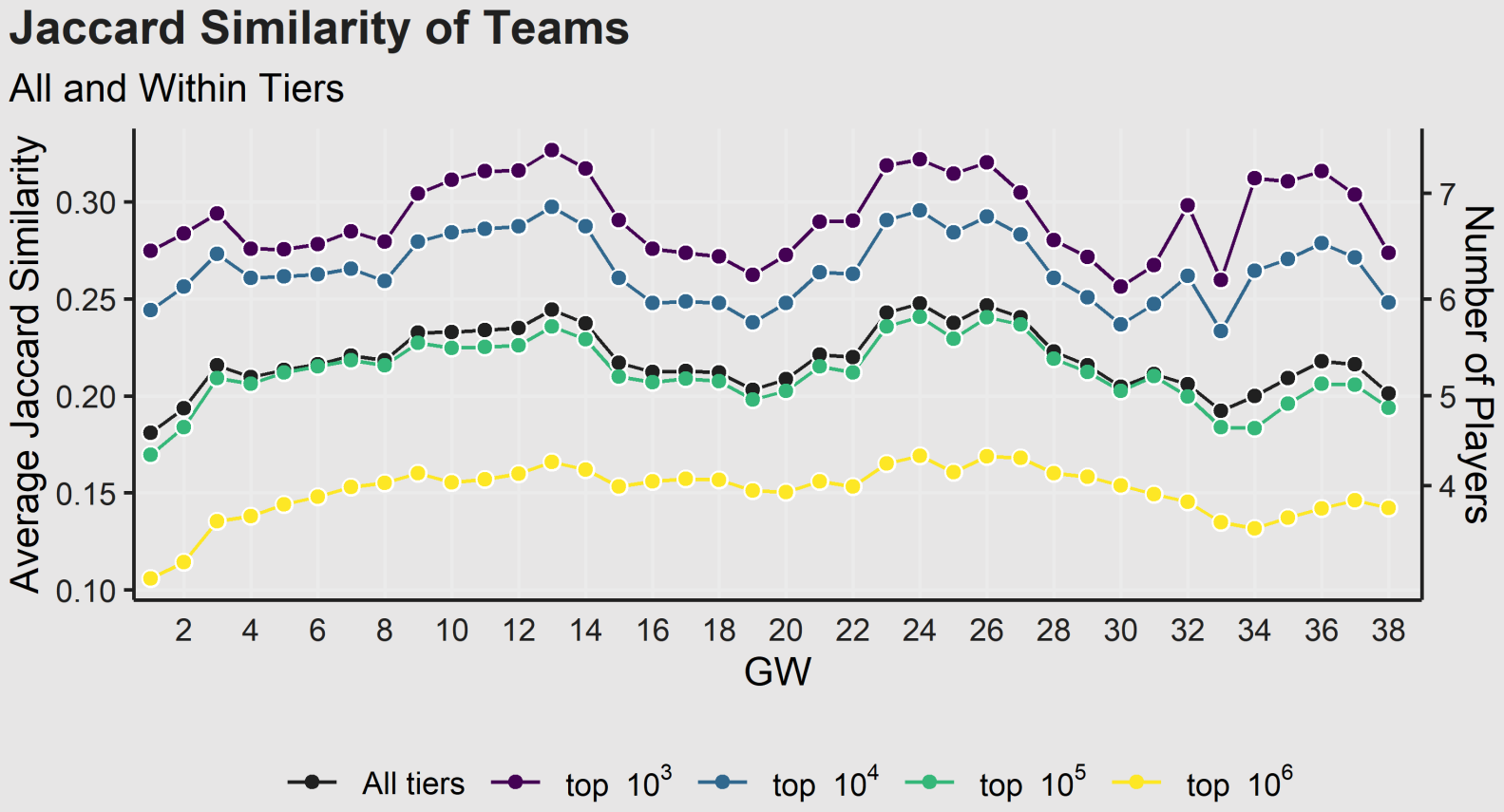
- From the network we can identify **clusters of players** based upon their section frequency.
- Find that four clusters can describe the different groups with **three of them** containing only **~30 players** (out of >600).



Template team

- We determine the similarity between two teams A and B through the **Jaccard Similarity** measure.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$



Conclusions -FPL

- There are consistently strong performing managers – much **more so than random** chance would suggest
- Clear patterns of **good decision making** + **identification of value players**
- **Network analysis** can inform us about the essential players
- Remarkable collective behaviour amongst managers – the **template team**

2. Snooker and its ranking problem

Motivation

- How to rank elements in a complex system?
- Ideally using a quantitative approach
- Requires detailed data of interactions
- Inherent networked representation

Motivation

- How to rank elements in a complex system?
- Ideally using a quantitative approach
- Requires detailed data of interactions
- Inherent networked representation

Motivation

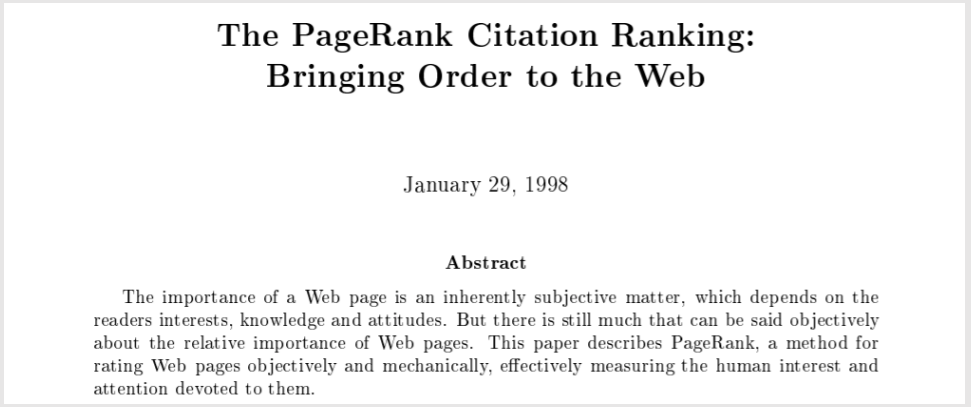
- How to rank elements in a complex system?
- Ideally using a quantitative approach
- Requires detailed data of interactions
- Inherent networked representation

Motivation

- How to rank elements in a complex system?
- Ideally using a quantitative approach
- Requires detailed data of interactions
- Inherent networked representation

Motivation

- How to rank elements in a complex system?
- Ideally using a quantitative approach
- Requires detailed data of interactions
- Inherent networked representation



The PageRank Citation Ranking:
Bringing Order to the Web

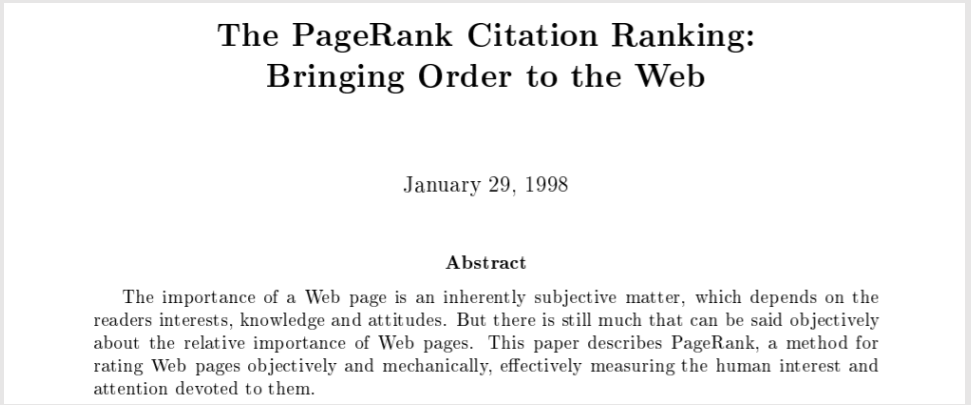
January 29, 1998

Abstract

The importance of a Web page is an inherently subjective matter, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them.

Motivation

- How to rank elements in a complex system?
- Ideally using a quantitative approach
- Requires detailed data of interactions
- Inherent networked representation

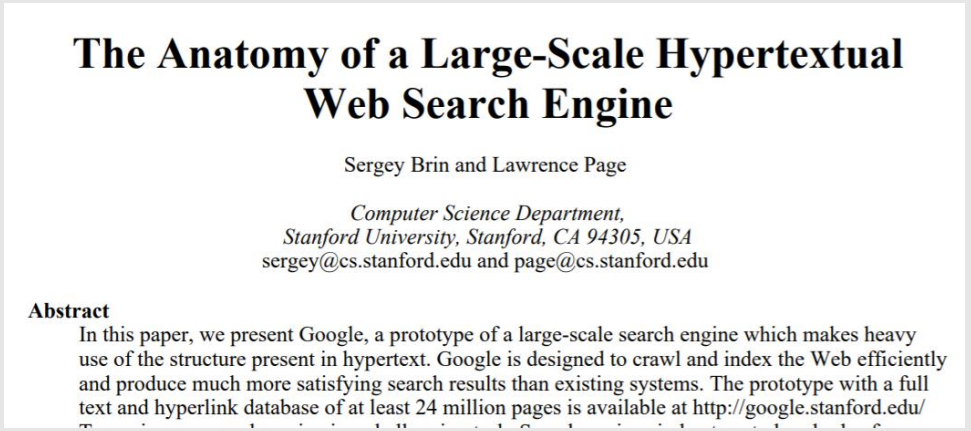


The PageRank Citation Ranking:
Bringing Order to the Web

January 29, 1998

Abstract

The importance of a Web page is an inherently subjective matter, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them.



The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Computer Science Department,
Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/

Motivation

- How to rank elements in a complex system?
- Ideally using a quantitative approach
- Requires detailed data of interactions
- Inherent networked representation

The PageRank Citation Ranking:
Bringing Order to the Web

January 29, 1998

Abstract

The importance of a Web page is an inherently subjective matter, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them.

The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Computer Science Department,
Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/

So… snooker?

So… snooker?

So… snooker?

So… snooker?

So… snooker?

So… snooker?

So… snooker?

- Each year there are a number of tournaments that the player's compete in
- Classically, ranked on a point basis obtained from results
- From 2013 instead based upon total prize money

So… snooker?

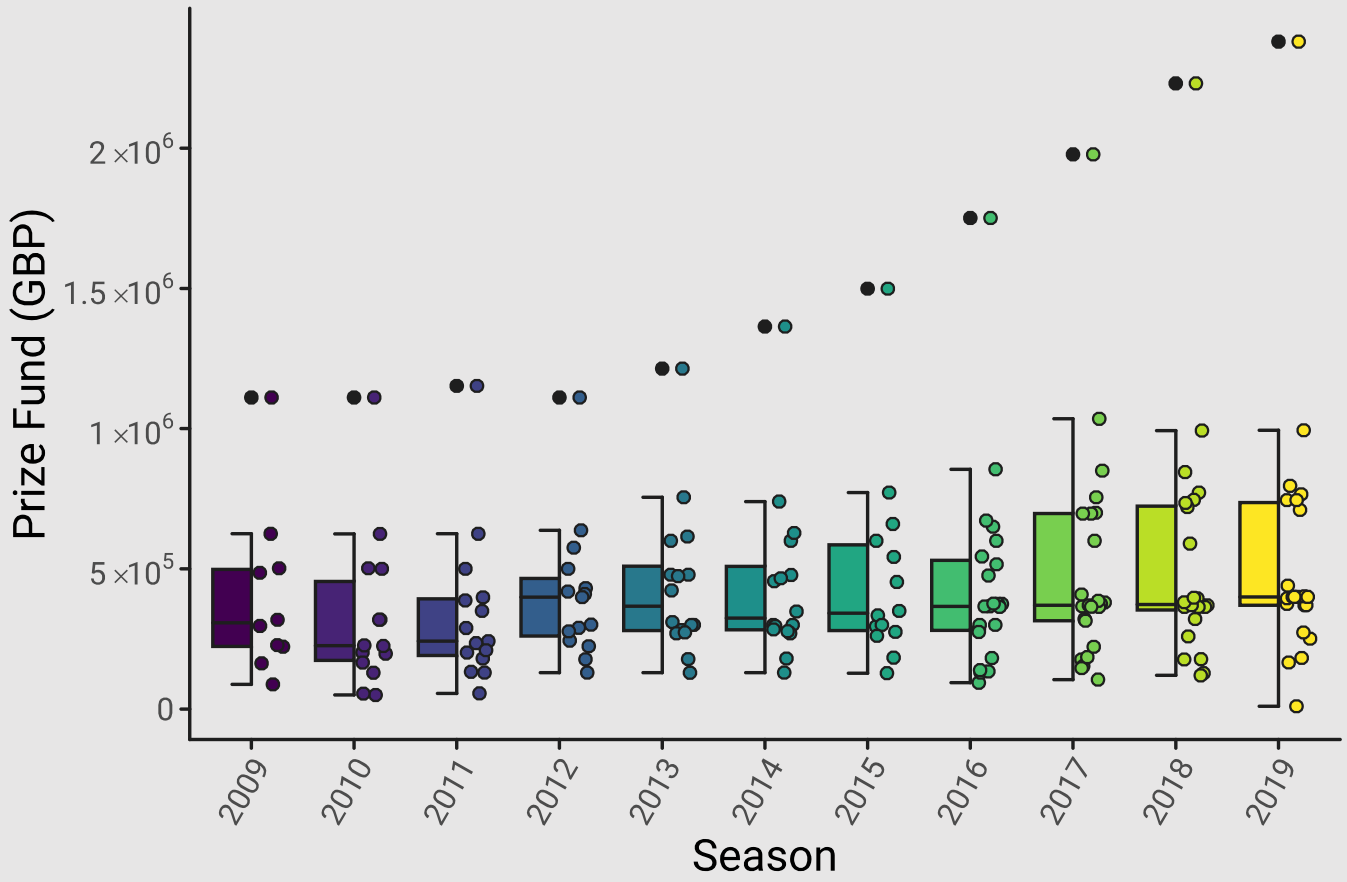
- Each year there are a number of tournaments that the player's compete in
- Classically, ranked on a point basis obtained from results
- From 2013 instead based upon total prize money

So… snooker?

- Each year there are a number of tournaments that the player's compete in
- Classically, ranked on a point basis obtained from results
- From 2013 instead based upon total prize money

So… snooker?

- Each year there are a number of tournaments that the player's compete in
- Classically, ranked on a point basis obtained from results
- From 2013 instead based upon total prize money



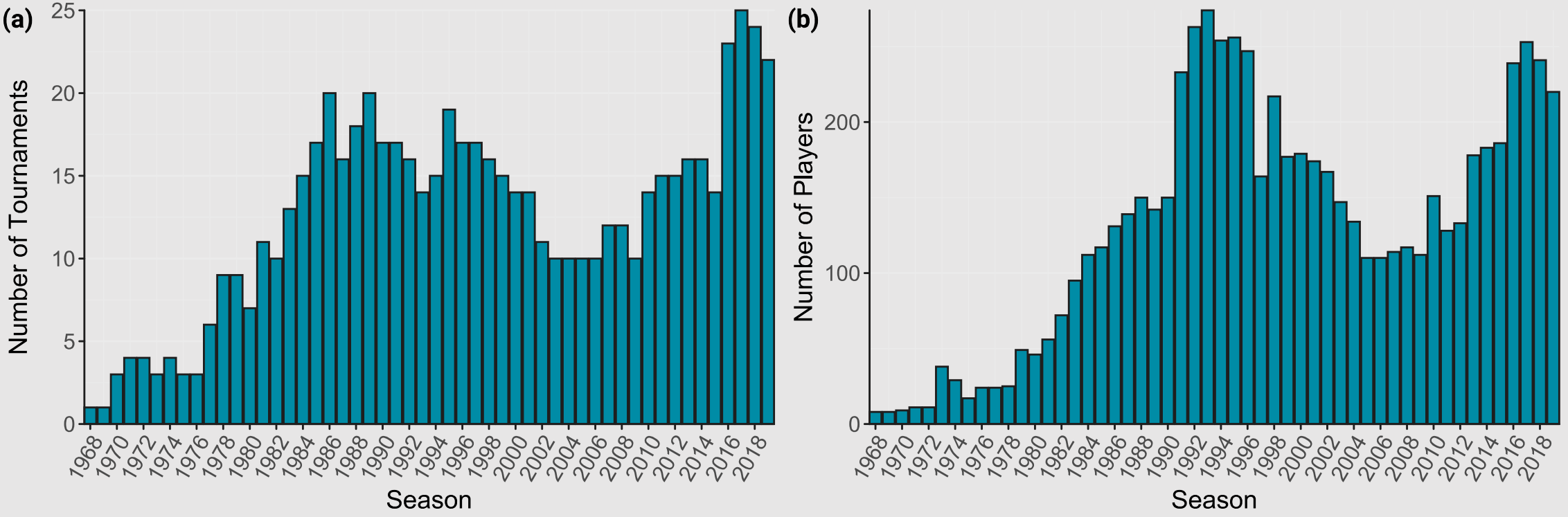
Data - collection

- Historical results from 1908-2020 stored on **cuetracker.net**

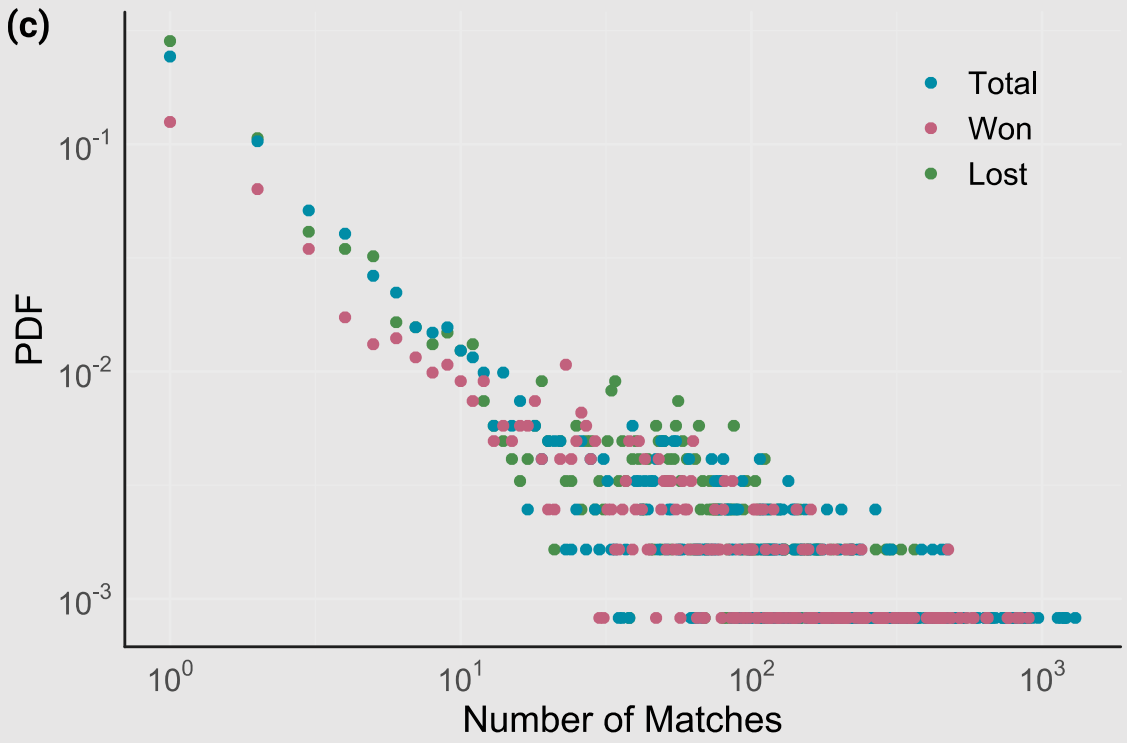
- Entire website scraped we focus on **professional era from 1968**

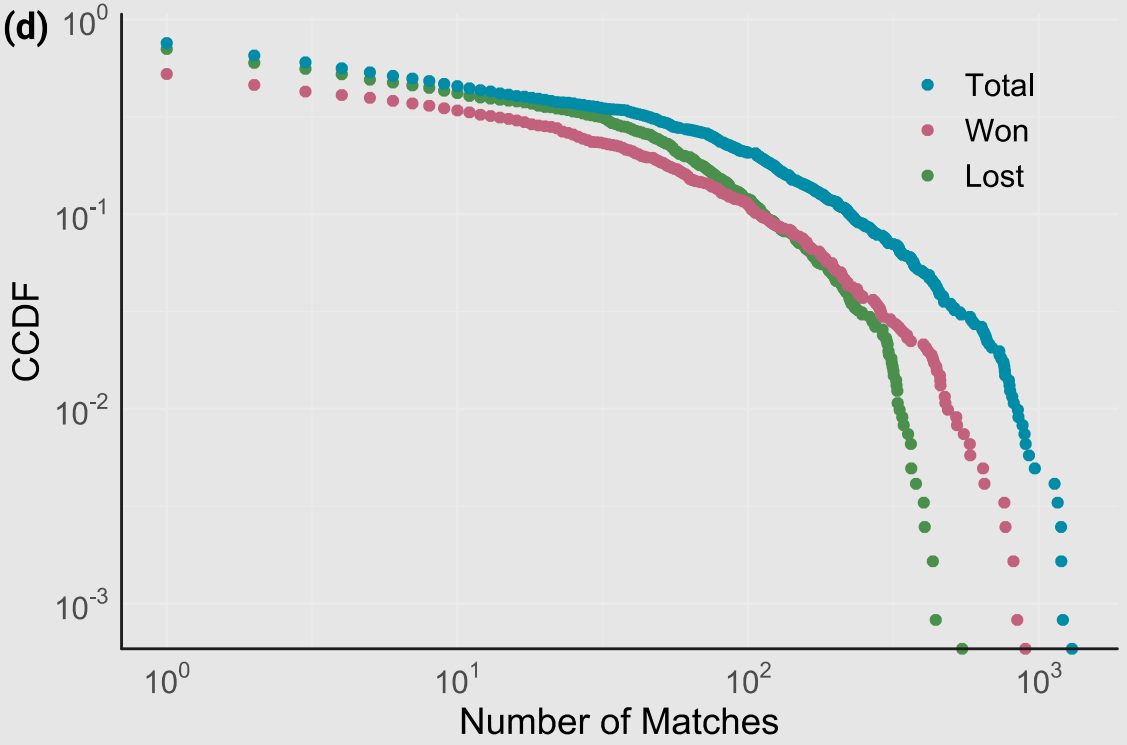
    - 657 tournaments
    - 1221 players
    - > 47,000 matches
    - Official rankings from 1975 onwards

Data – initial analysis

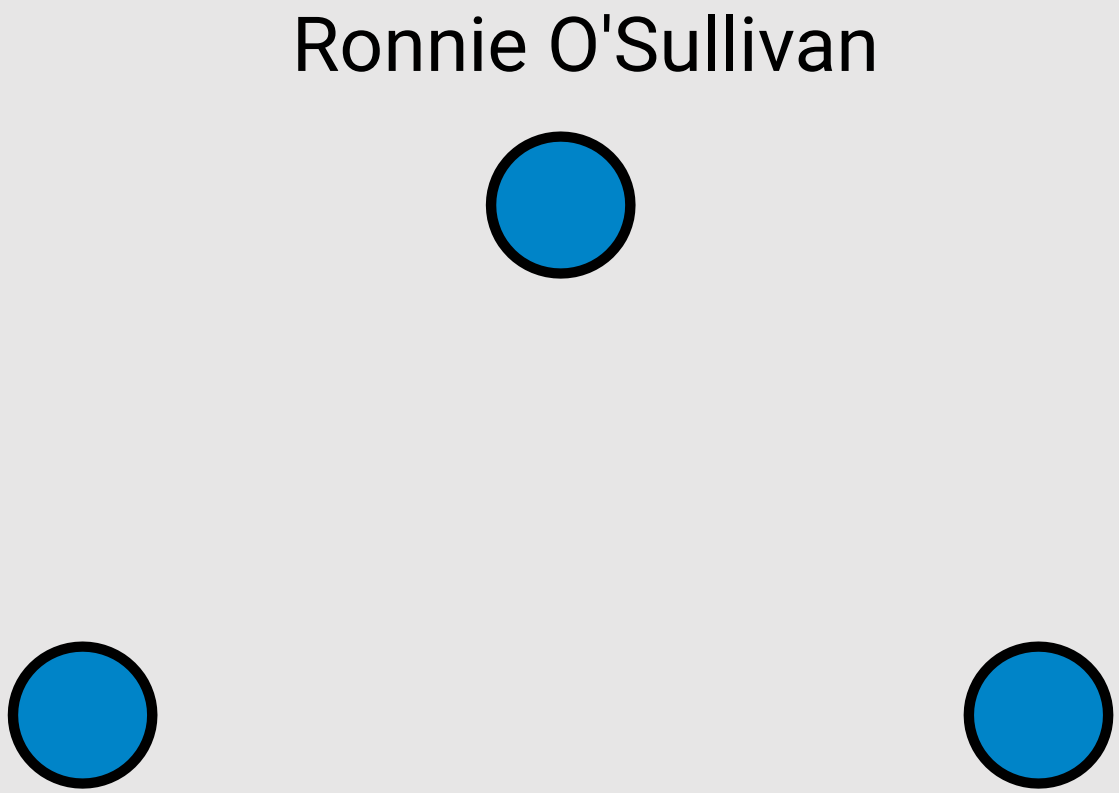


Data – initial analysis

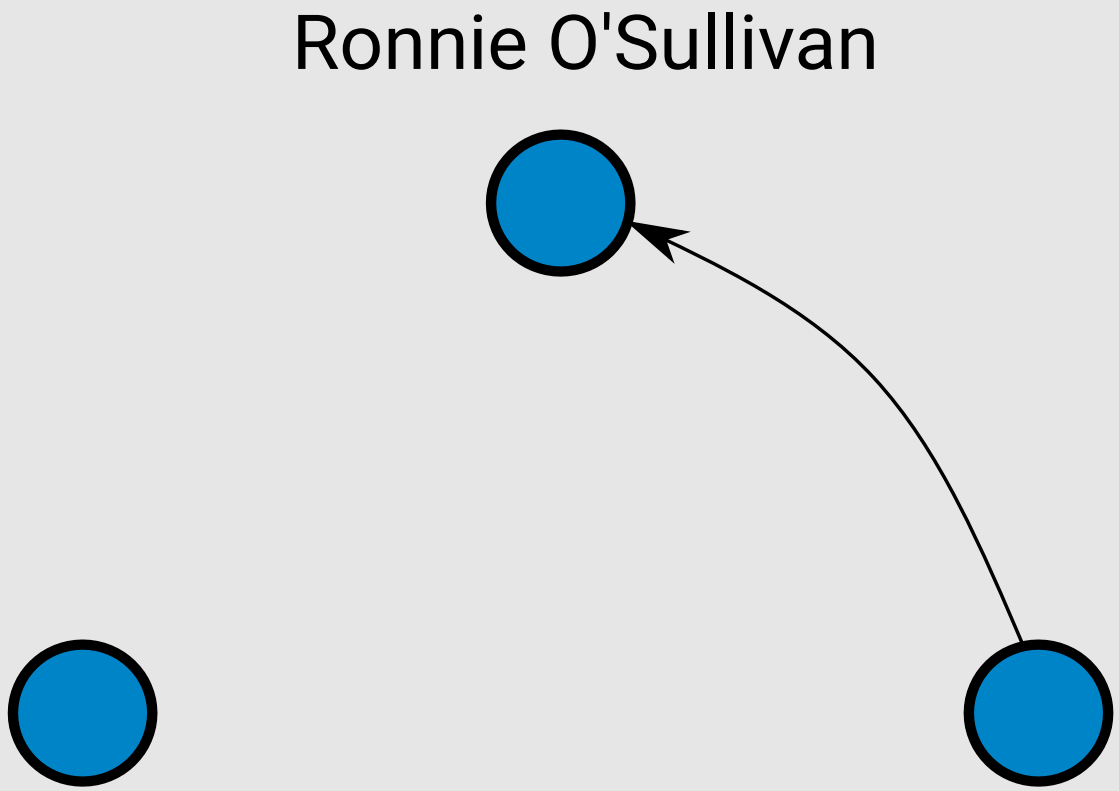




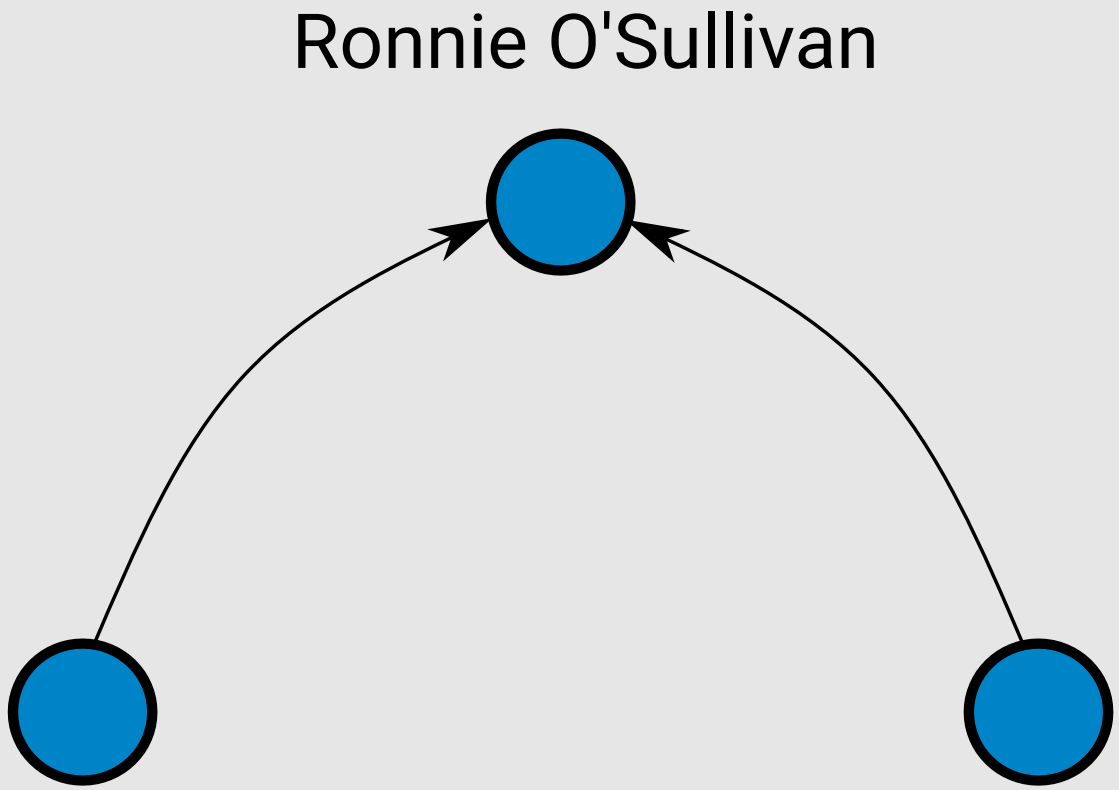
Snooker’s complex network



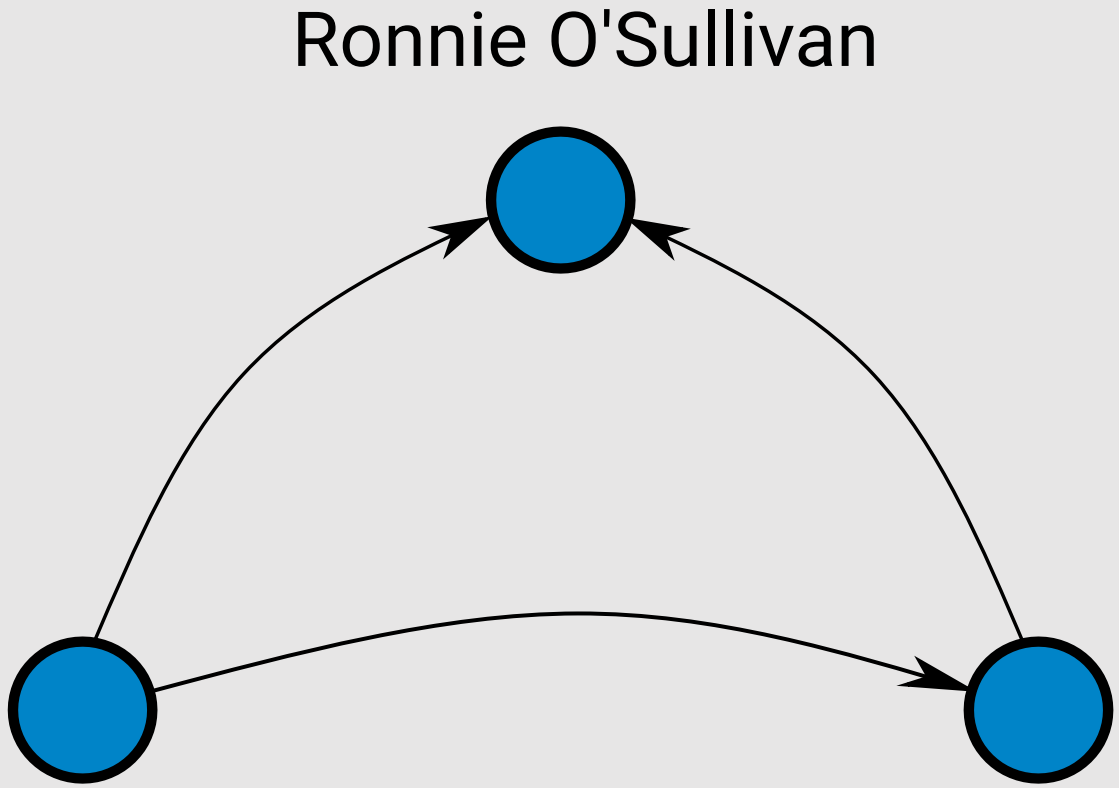
Snooker’s complex network



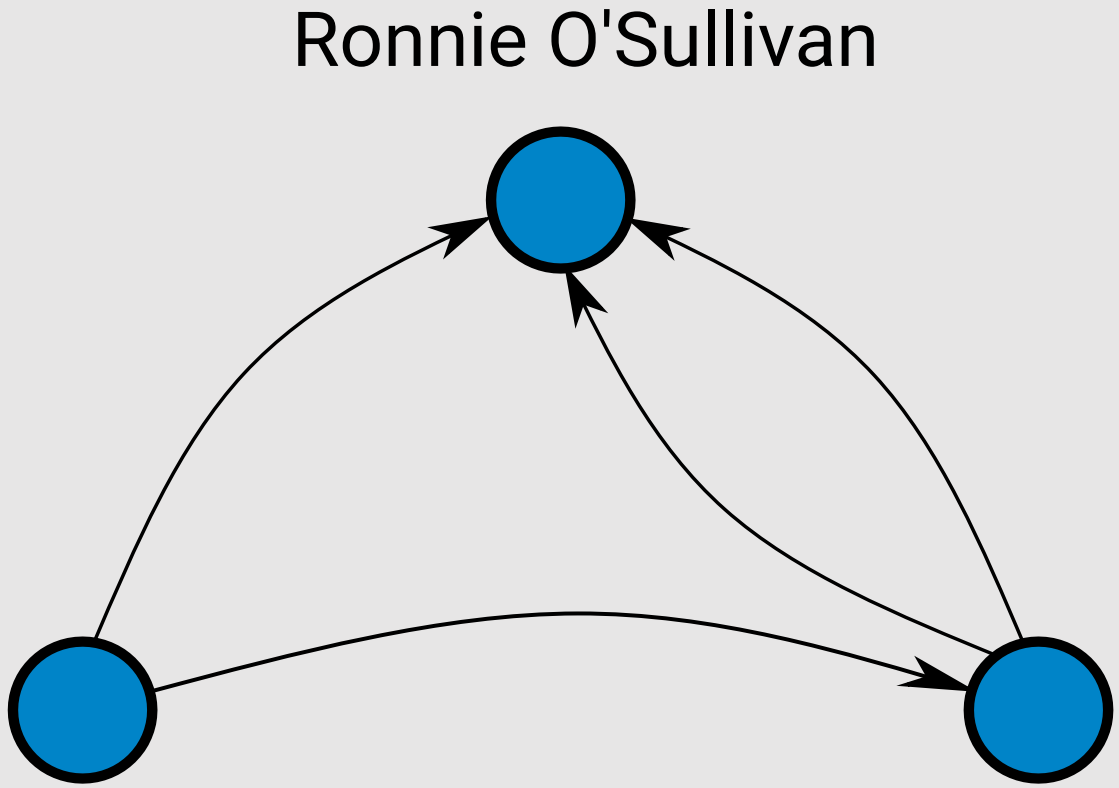
Snooker’s complex network



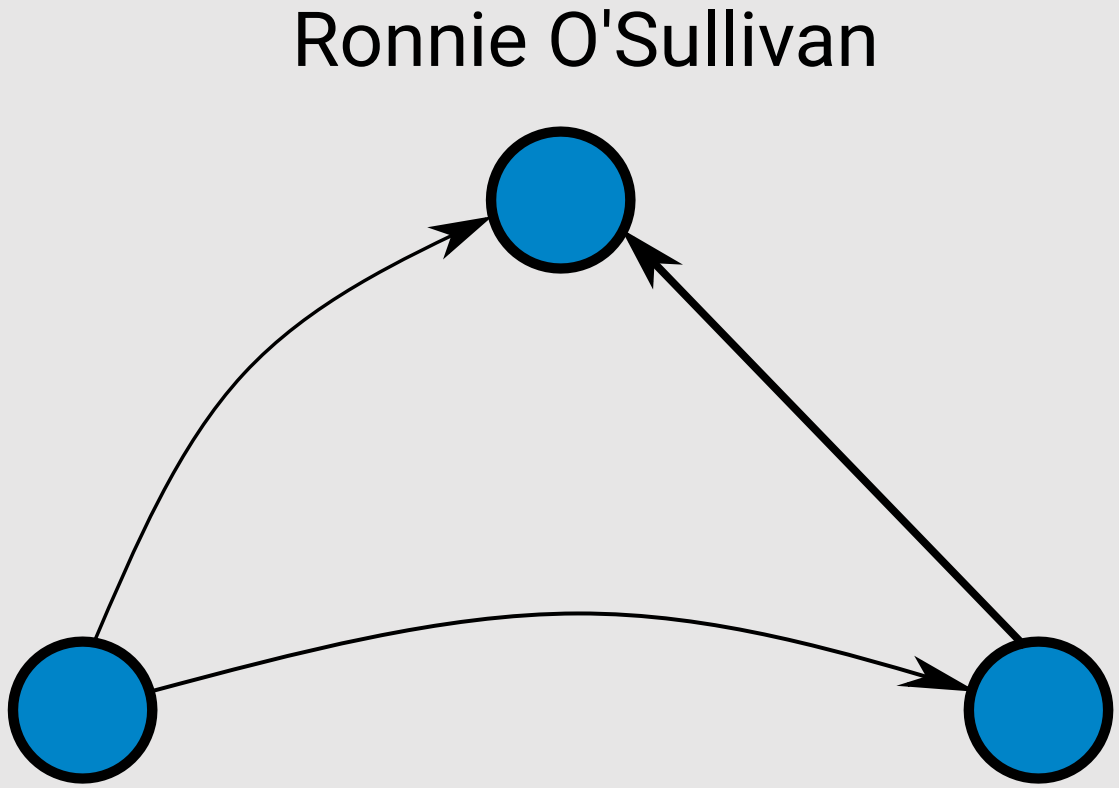
Snooker’s complex network



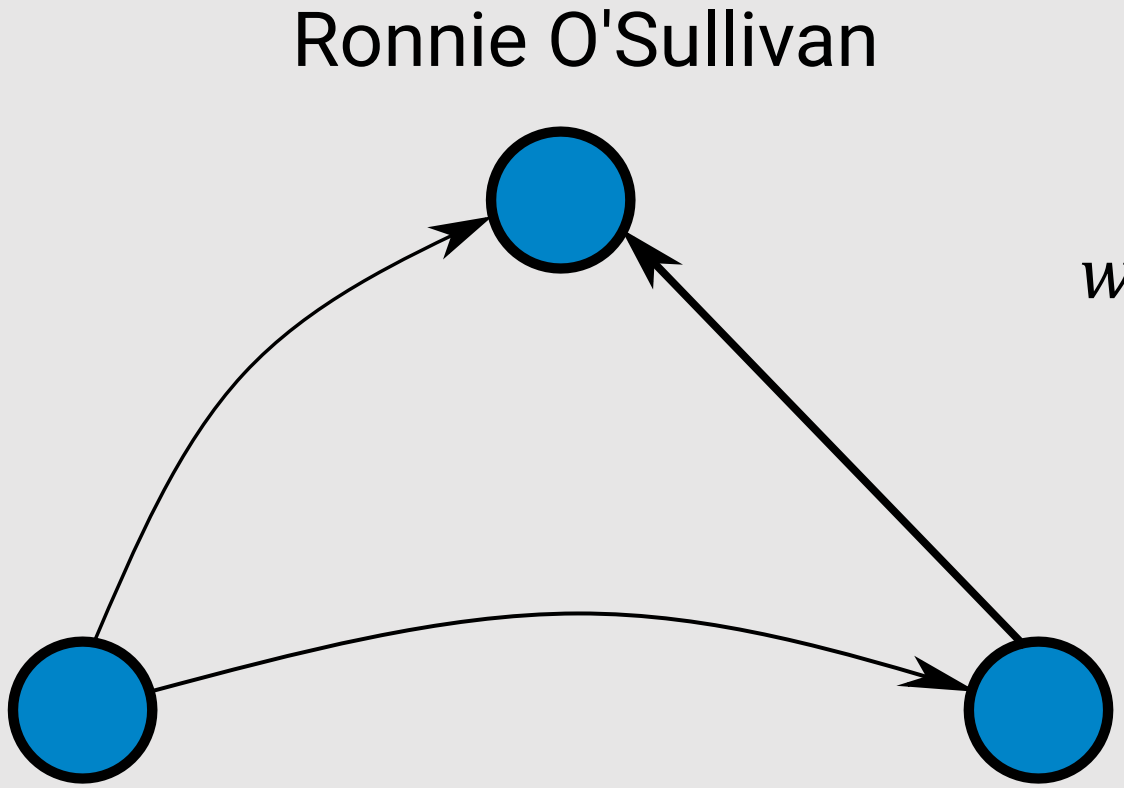
Snooker’s complex network



Snooker’s complex network



Snooker's complex network



w_{ji} No. of times player j has lost to player i

Mathematical framework

- Prestige score for player i, P_i, obtained from N coupled equations

$$P_i = (1-q)\sum_j P_j \frac{w_{ji}}{k_j^{out}} + \frac{q}{N} + \frac{1-q}{N}\sum_j P_j \delta(k_j^{out})$$

Mathematical framework

- Prestige score for player i, P_i, obtained from N coupled equations

$$P_i = (1-q)\sum_j P_j \frac{w_{ji}}{k_j^{out}} + \frac{q}{N} + \frac{1-q}{N}\sum_j P_j \delta(k_j^{out})$$

Player i's level of prestige

Mathematical framework

- Prestige score for player i, P_i, obtained from N coupled equations

Damping factor for unif. redistribution of prestige

$$P_i = (1-q)\sum_j P_j \frac{w_{ji}}{k_j^{out}} + \frac{q}{N} + \frac{1-q}{N}\sum_j P_j \delta(k_j^{out})$$

Player i's level of prestige

Mathematical framework

- Prestige score for player i, P_i, obtained from N coupled equations

No. of times player j has lost to player i

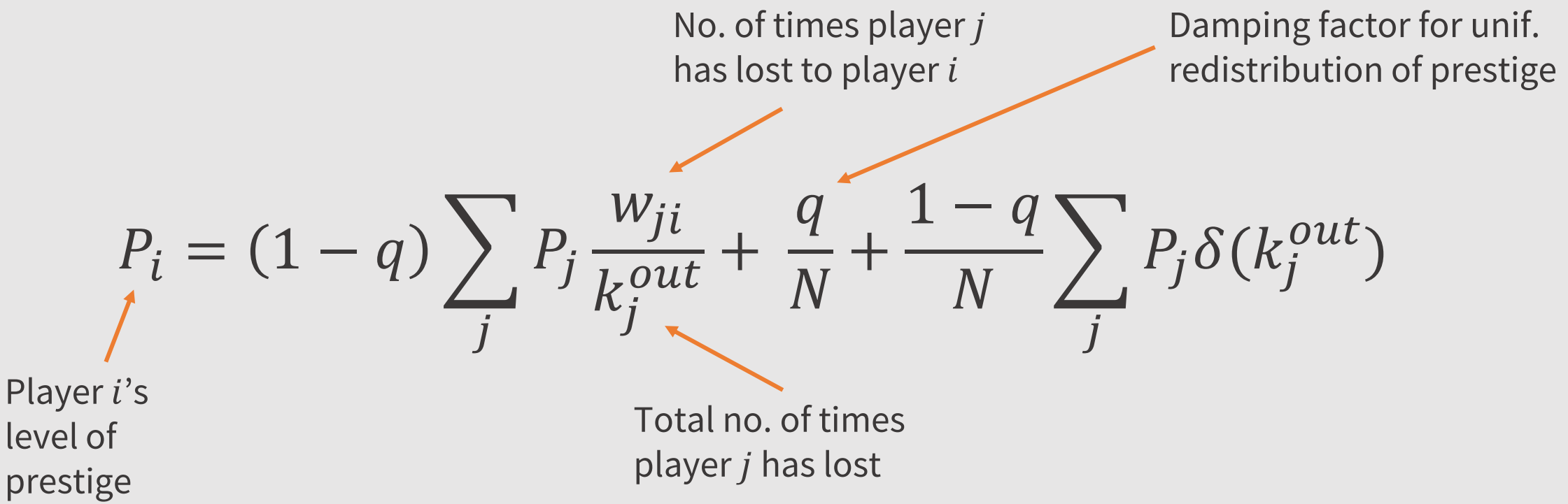
Damping factor for unif. redistribution of prestige

$$P_i = (1-q)\sum_j P_j \frac{w_{ji}}{k_j^{out}} + \frac{q}{N} + \frac{1-q}{N}\sum_j P_j \delta(k_j^{out})$$

Player i's level of prestige

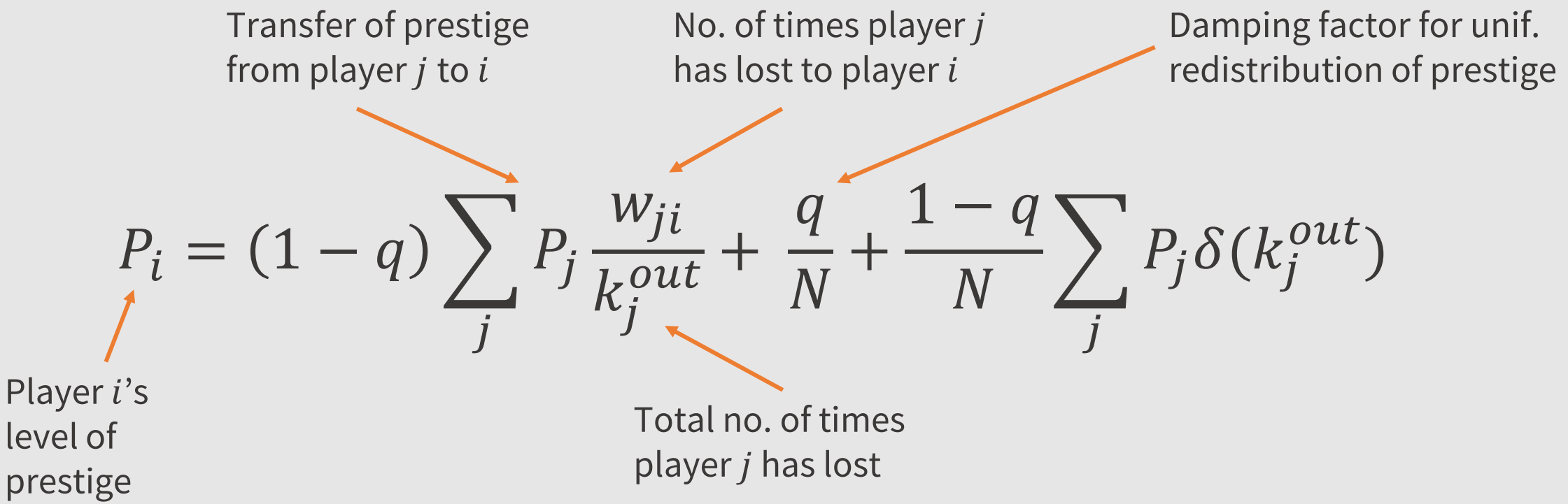
Mathematical framework

- Prestige score for player i, P_i, obtained from N coupled equations



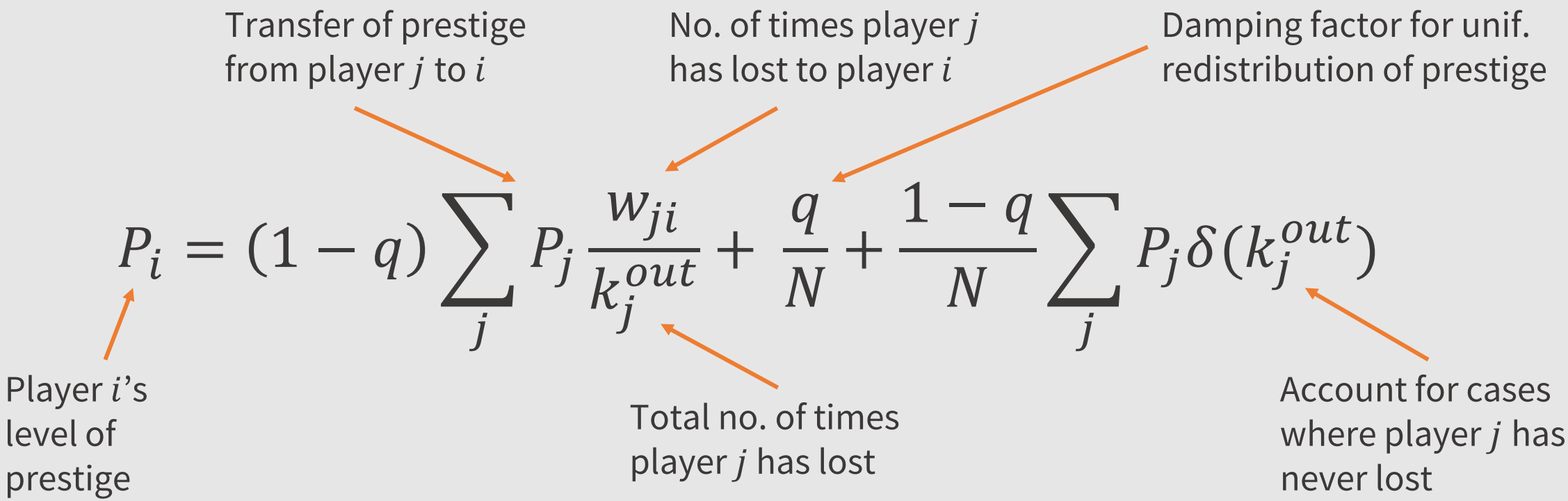
Mathematical framework

- Prestige score for player i, P_i, obtained from N coupled equations



Mathematical framework

- Prestige score for player i, P_i, obtained from N coupled equations



Rankings – all-time

The Top 10 Snooker Players

1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players
1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players
1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players

1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
4	Stephen Hendry	0.0164	
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players

1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
4	Stephen Hendry	0.0164	
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players

1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
3	Mark Williams	0.0169	
4	Stephen Hendry	0.0164	
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players

1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
3	Mark Williams	0.0169	
4	Stephen Hendry	0.0164	
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players

1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
2	Ronnie O'Sullivan	0.0201	
3	Mark Williams	0.0169	
4	Stephen Hendry	0.0164	
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players

1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
2	Ronnie O'Sullivan	0.0201	
3	Mark Williams	0.0169	
4	Stephen Hendry	0.0164	
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players

1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
1	John Higgins	0.0204	
2	Ronnie O'Sullivan	0.0201	
3	Mark Williams	0.0169	
4	Stephen Hendry	0.0164	
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

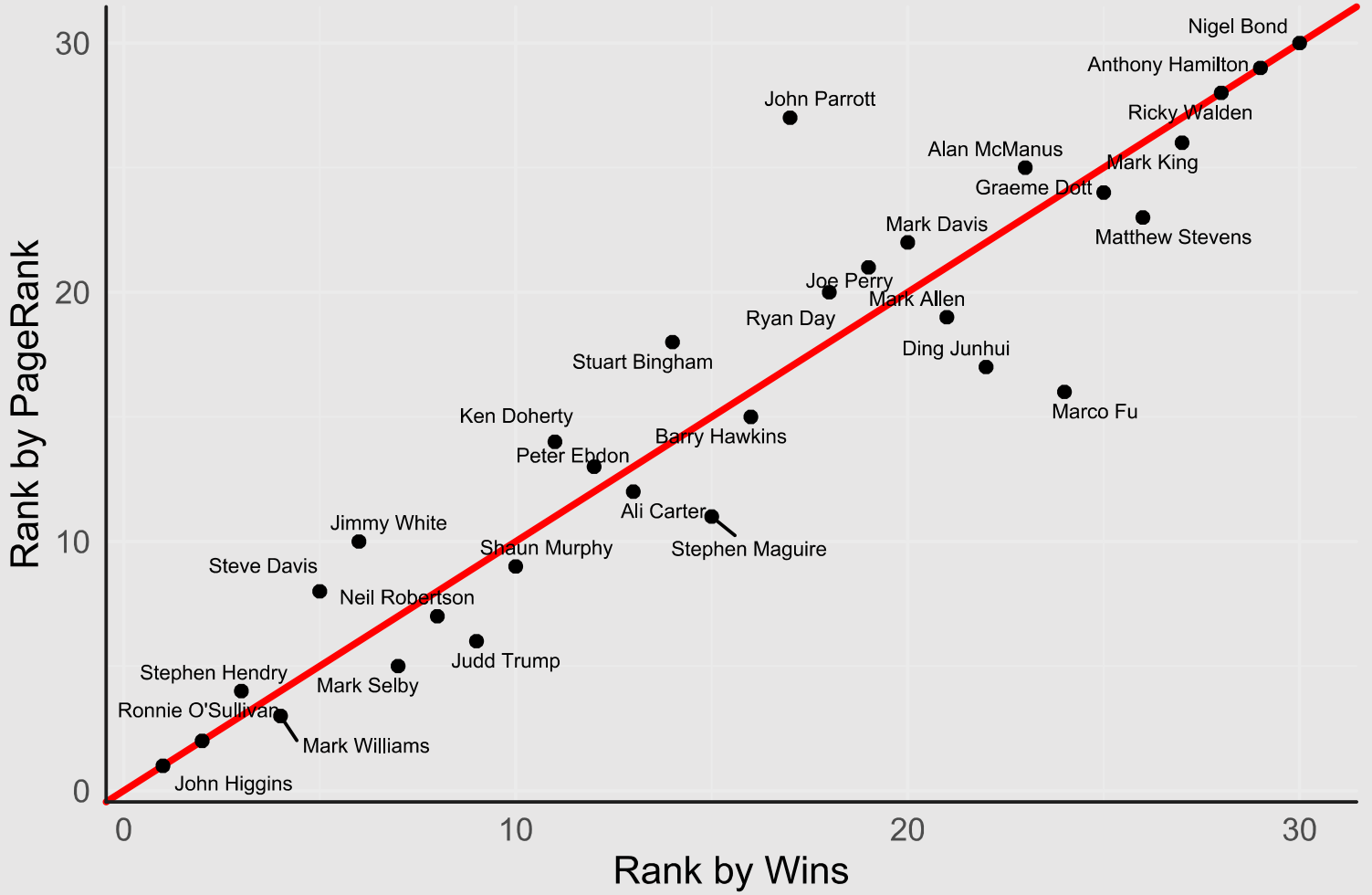
[1] Calculated via a complex networks approach.

Rankings – all-time

The Top 10 Snooker Players
1968 - 2020

RANK[1]	PLAYER	SCORE	NATION
1	John Higgins	0.0204	
2	Ronnie O'Sullivan	0.0201	
3	Mark Williams	0.0169	
4	Stephen Hendry	0.0164	
5	Mark Selby	0.0149	
6	Judd Trump	0.0136	
7	Neil Robertson	0.0134	
8	Steve Davis	0.0129	
9	Shaun Murphy	0.0126	
10	Jimmy White	0.0116	

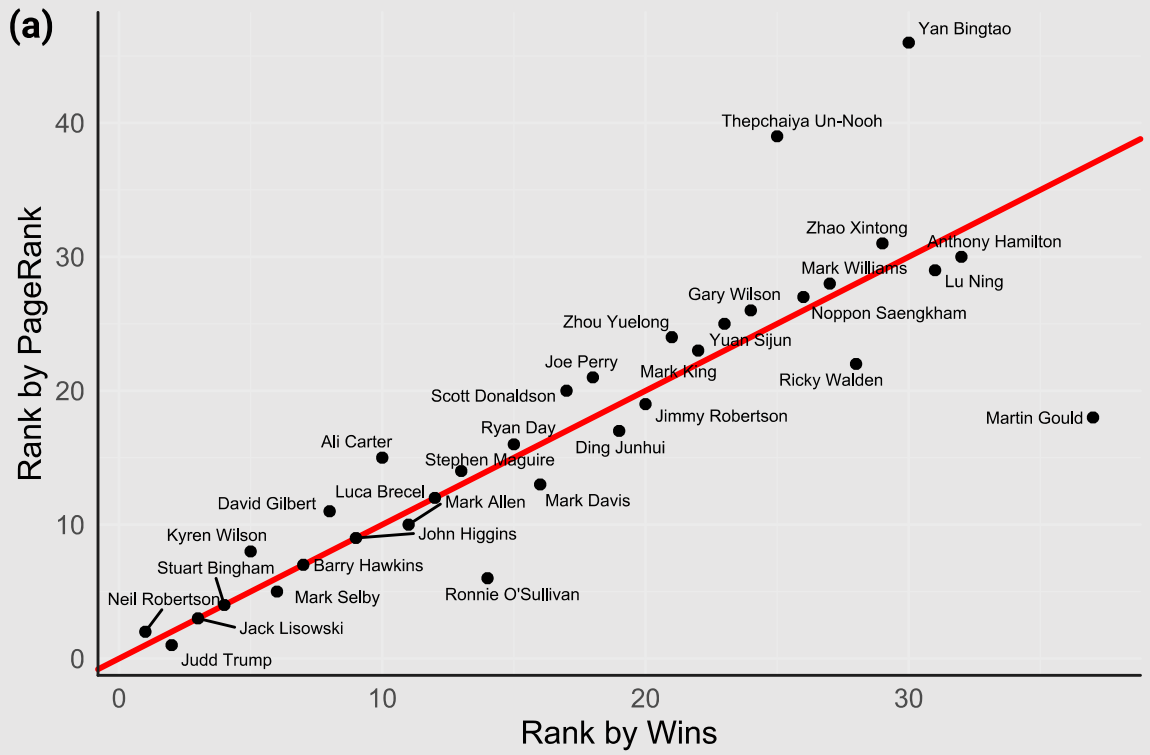
[1] Calculated via a complex networks approach.

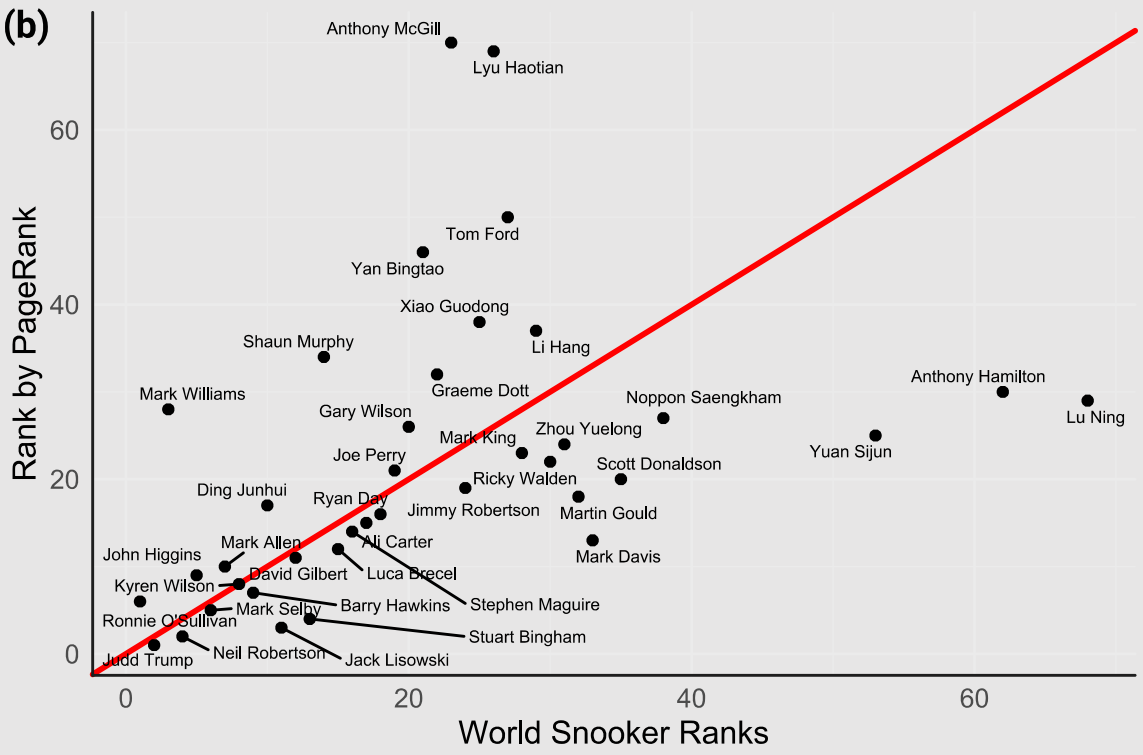


@obrienj_

Rankings – temporal periods

- Any time period may be considered by filtering the matches

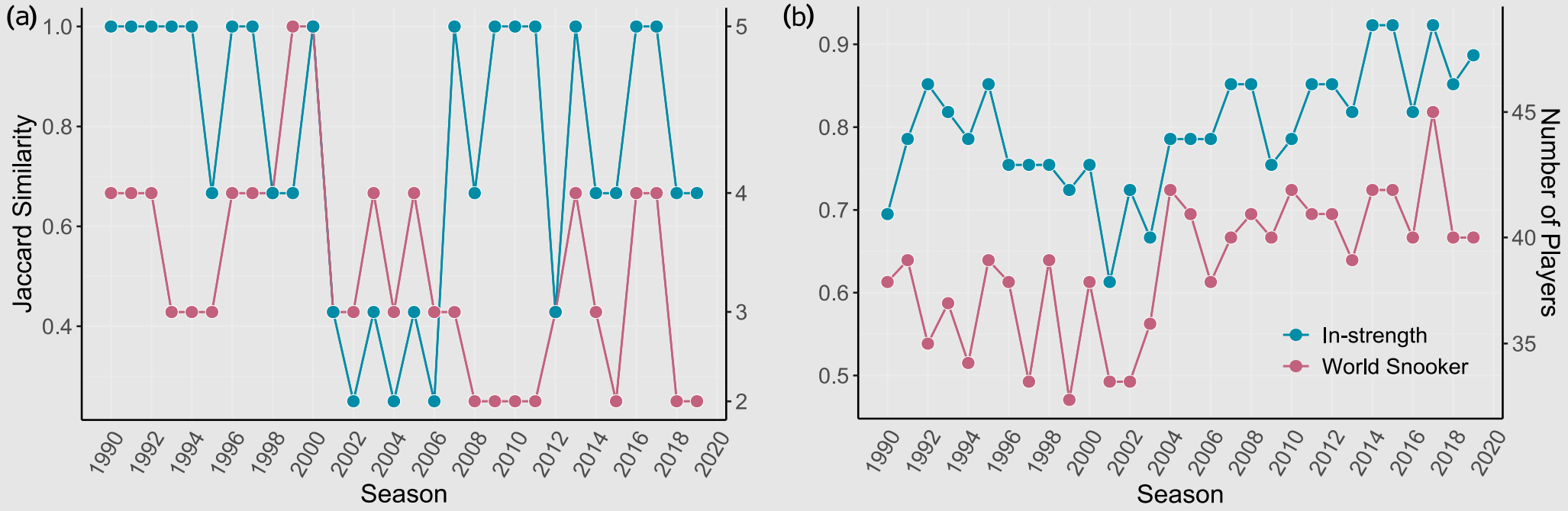




Rankings – Similarity of approaches

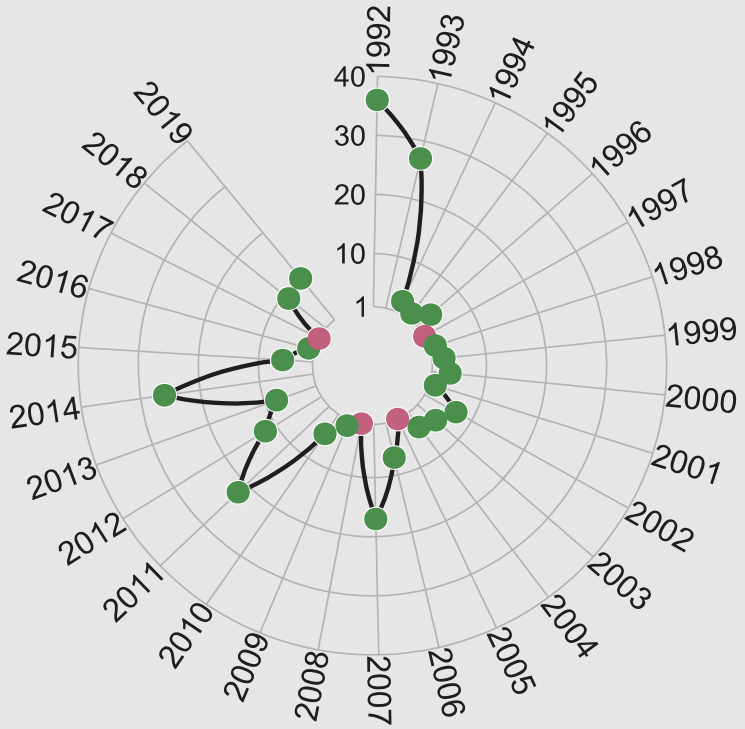
- Two rankings from approaches A and B the Jaccard similarity is

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

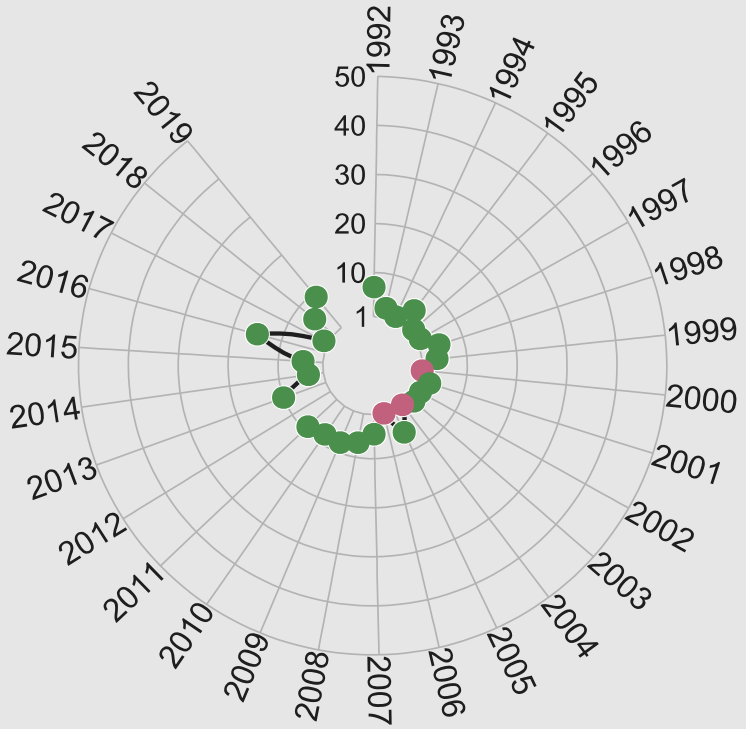


Visualizations – Rank Clocks

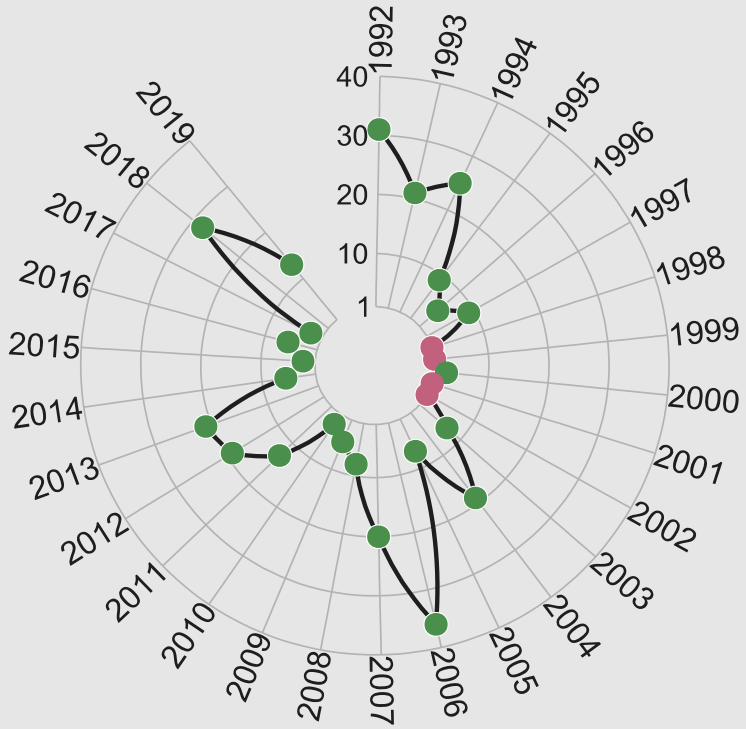
John Higgins



Ronnie O'Sullivan



Mark Williams



Conclusions - Snooker

- Ranking is difficult – **networks** can help
- **Detailed data** of competitions is required
- Our approach allows **detailed inferences** to be made including **all-time ranks**, arbitrary **temporal periods**, …
- Captures **stronger features** of other approaches while also considering the **quality of opponent** defeated

Collaborators

James Gleeson

David O'Sullivan

University of Limerick

Thanks

PLOS ONE

RESEARCH ARTICLE

Identification of skill in an online game: The case of Fantasy Premier League

Joseph D. O'Brien*, James P. Gleeson, David J. P. O'Sullivan

MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland

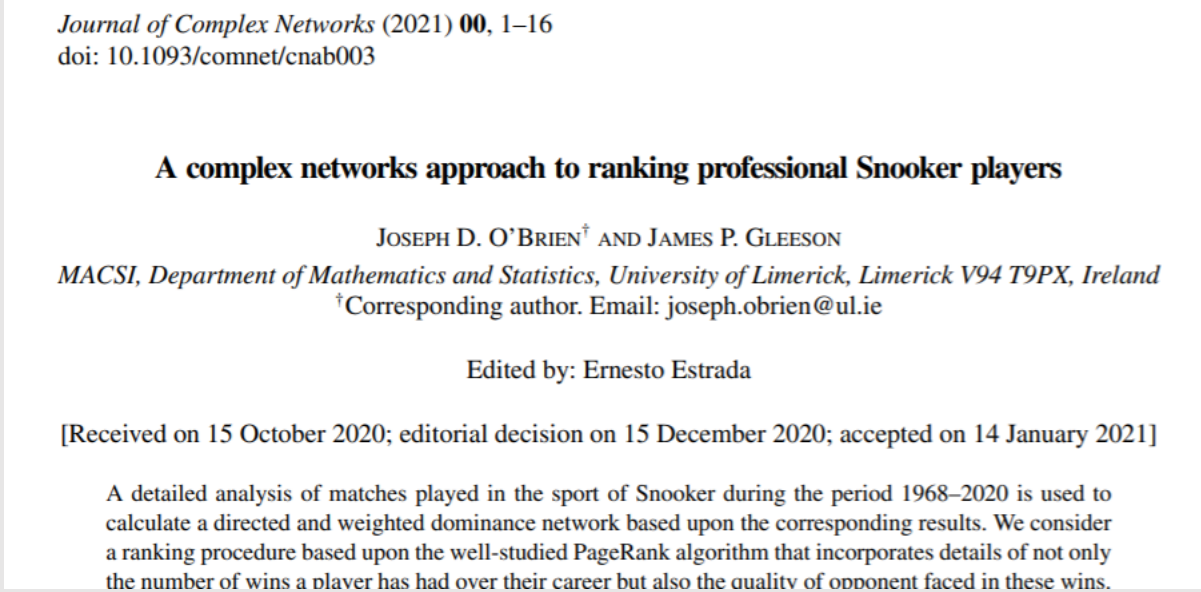
* joseph.obrien@ul.ie

Check for updates

Abstract

In all competitions where results are based upon an individual's performance the question of whether the outcome is a consequence of skill or luck arises. We explore this question through an analysis of a large dataset of approximately one million contestants playing *Fantasy Premier League*, an online fantasy sport where managers choose players from the

PLOS ONE 16(3): e0246698.



Journal of Complex Networks (2021) **00**, 1–16
doi: 10.1093/comnet/cnab003

A complex networks approach to ranking professional Snooker players

JOSEPH D. O'BRIEN† AND JAMES P. GLEESON

MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick V94 T9PX, Ireland

†Corresponding author. Email: joseph.obrien@ul.ie

Edited by: Ernesto Estrada

[Received on 15 October 2020; editorial decision on 15 December 2020; accepted on 14 January 2021]

A detailed analysis of matches played in the sport of Snooker during the period 1968–2020 is used to calculate a directed and weighted dominance network based upon the corresponding results. We consider a ranking procedure based upon the well-studied PageRank algorithm that incorporates details of not only the number of wins a player has had over their career but also the quality of opponent faced in these wins.

J. of Complex Networks 8(6), cnab003

obrienjoey/snooker_rankings

Fondúireacht Eolaíochta Éireann
Dá bhfuil romhainn
Science Foundation Ireland
For what's next

CIT Research Methods Seminar

Thank you for listening!

✉ joseph.obrien@ul.ie

@obrienj_