UK Data Service

Social Network Analysis: Fundamental Concepts

Dr Diarmuid McDonnell
UK Data Service
01 September 2020

New Forms of Data Training Series

Upcoming webinars:

- Social Network Analysis: Getting and Marshalling Data (15 September 2020)
- Social Network Analysis: Techniques and Methods of Analysis (29 September 2020)

Upcoming coding demonstrations:

- Text Mining in Python (02 – 30 September 2020)

Past webinars:

- Text-Mining: Advanced Options
- Text-Mining: Basic Processes

UK Data Service

Table of Contents

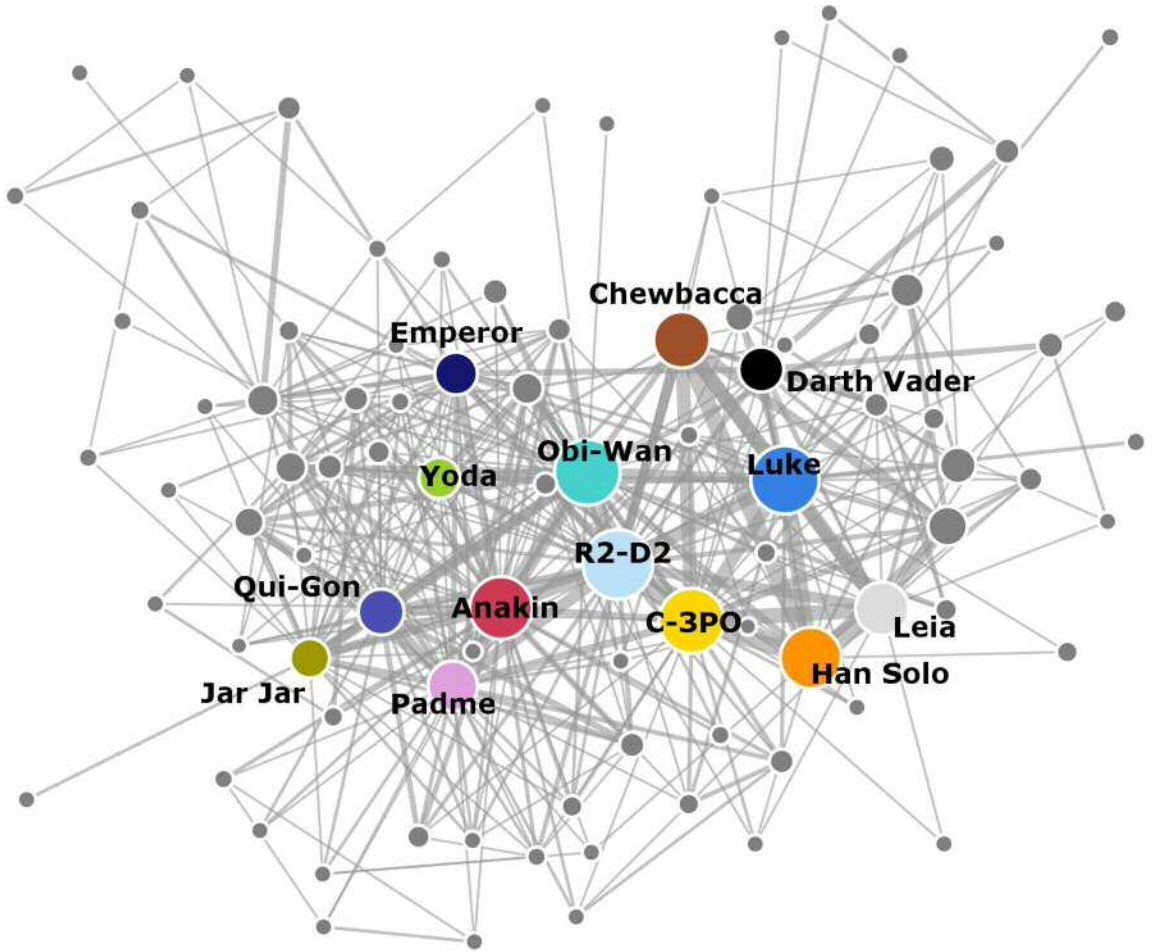
UK Data Service

Why this training series?

Many who have seen the potential offered by network analysis have found it difficult to come to grips with the highly technical and mathematical language that necessarily characterises much of the discussion in the technical literature.

(Scott, 2017: 3)

In a social network far, far away…



Source: http://evelinag.com/blog/2015/12-15-star-wars-social-network/

UK Data Service

Overview

UK Data Service

What is Social Network Analysis?

Social Network Analysis (SNA) is a methodological and conceptual toolbox for the measurement, systematic description, and analysis of patterns in relational structures in the social world (Caiani, 2014).

A relation is a distinctive type of connection or tie between two entities (Wasserman & Faust, 1994).

Relations are the building blocks of networks, and thus SNA is concerned with and most appropriate for analyses of data capturing relations between units of analysis (Scott, 2017).

Why should you consider SNA for your research?

A social network is the phenomenon you are trying to describe and explain (*Y*).

Features or properties of a social network help you describe and explain a different phenomenon (*X*).

> *illness, disability, health behaviors, health care use, and death in one person are associated with similar outcomes in numerous others to whom that person is tied, and there can be a nonbiological transmission of illness.* (Smith & Christakis, 2008: 420)

When should you use SNA?

When you are dealing with **relational data** i.e., data capturing relationships and connections between units of analysis.

This is in contrast to **attributional data**, which captures the attributes – characteristics, demographics etc – of your units of analysis.

When should you use SNA?

name	sex	age	employed
John	male	52	yes
Joan	female	45	yes
Jenny	female	25	no
Juliet	female	67	yes
Jack	male	19	no

	John	Joan	Jenny	Juliet	Jack
John	--	Friend	Colleague	Stranger	Stranger
Joan	Friend	--	Stranger	Wife	Stranger
Jenny	Colleague	Stranger	--	Friend	Stranger
Juliet	Stranger	Wife	Friend	--	Stranger
Jack	Stranger	Stranger	Stranger	Stranger	--

UK Data Service

What does SNA involve?

Scott (2017):

- Identifying and visualising patterns of relations between units of analysis.
- Examining structural properties/characteristics of these relations.
- Analysing implications of these relations on outcomes experienced by units of analysis.

As a result of its focus on the relational characteristics of the units of analysis, SNA requires distinctive data structures, methods of analysis and data visualisation techniques (Caiani, 2014).

How do you implement SNA in your research?

Hanneman & Riddle (2005), Owen-Smith (2017):

1. Pose a carefully articulated research question that requires understanding and/or analysis of a network.
2. Decide which units of analysis and types of relations to analyse i.e., who is connected and which relationships matter?
3. Collect or select a data set that provides relational data on your units of analysis.
4. Summarise the network and its key features using appropriate measures e.g., network size, density, cohesion etc.

Key concepts

UK Data Service

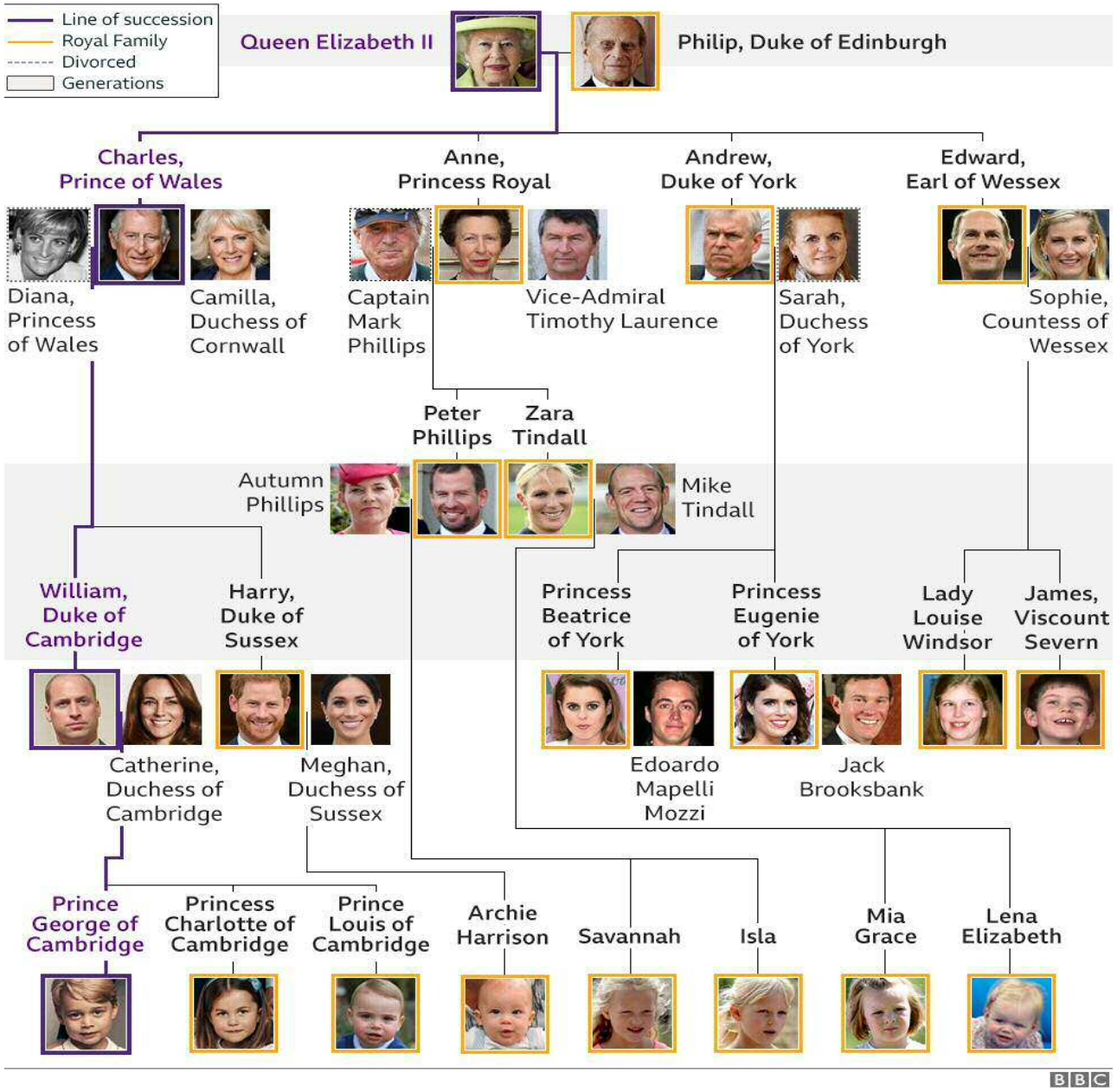
Networks in a nutshell

A network – whether social, physical, biological etc – is constructed from two main building blocks (Owen-Smith, 2017):

1. The **entities** that are (or can be) connected in a network.
2. The **connections** that exist (or could exist) between entities.

Therefore a network is an aggregation or collection of these entities and their connections.

For example, a family tree is a network containing individuals (**entities**) that are related through some type of familial tie (**connection**).



Source: https://www.bbc.com/news/uk-23272491

UK Data Service

Entities

The entities included in a network are known as **nodes**.

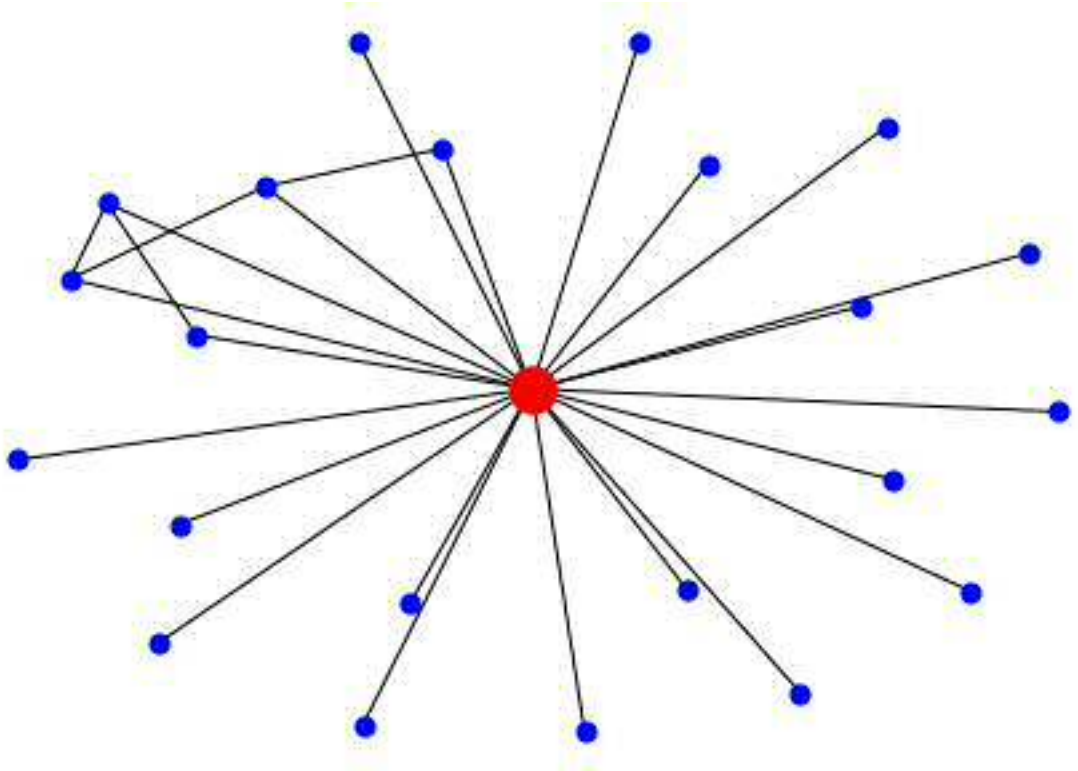
Nodes can be individuals, organisations, countries, animals, events, computers, train stations etc.

Nodes of particular interest are known as **ego** or **focal nodes**; nodes that are or could be connected to an ego are referred to as **alters**.

Two nodes that are or could be connected are called a **dyad**, while three nodes that are or could be connected are called a **triad**.

Entities

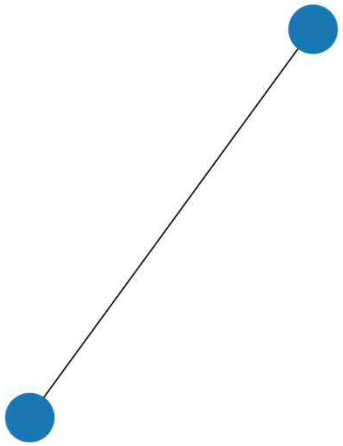
Ego Network



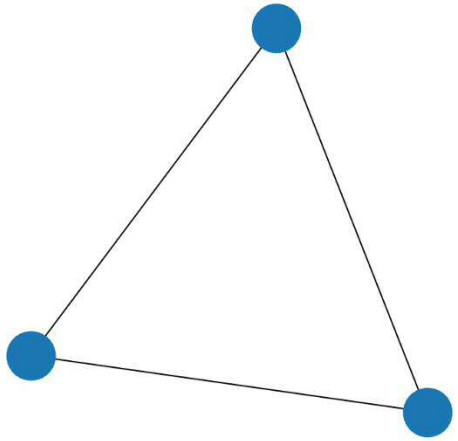
UK Data Service

Entities

Dyad



Triad



UK Data Service

Connections

Connections or relations between entities are known as **ties**.

There are a multitude of different types of ties present in the social world e.g., family relations, friendships, event attendance, club memberships, communal living, collegial etc.

It is possible for two entities to be connected by many different types of ties.

Important to acknowledge that your data will only capture a sample of **all** possible ties that exist between your nodes.

Connections

Ties have two dimensions:

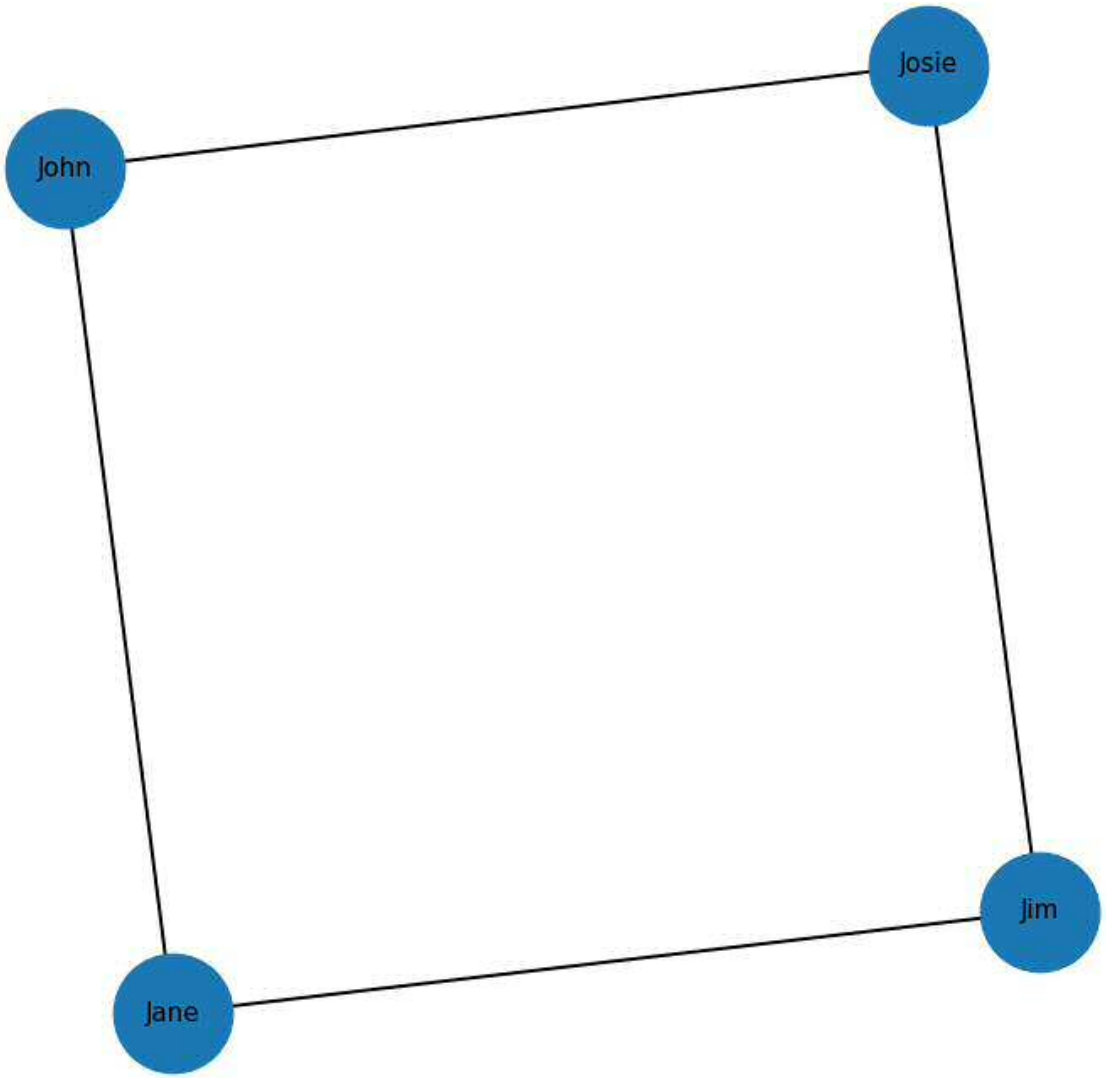
1. **Numeration / Strength**
   - Binary
   - Valued

2. **Directionality**
   - Undirected
   - Directed

UK Data Service

Connections

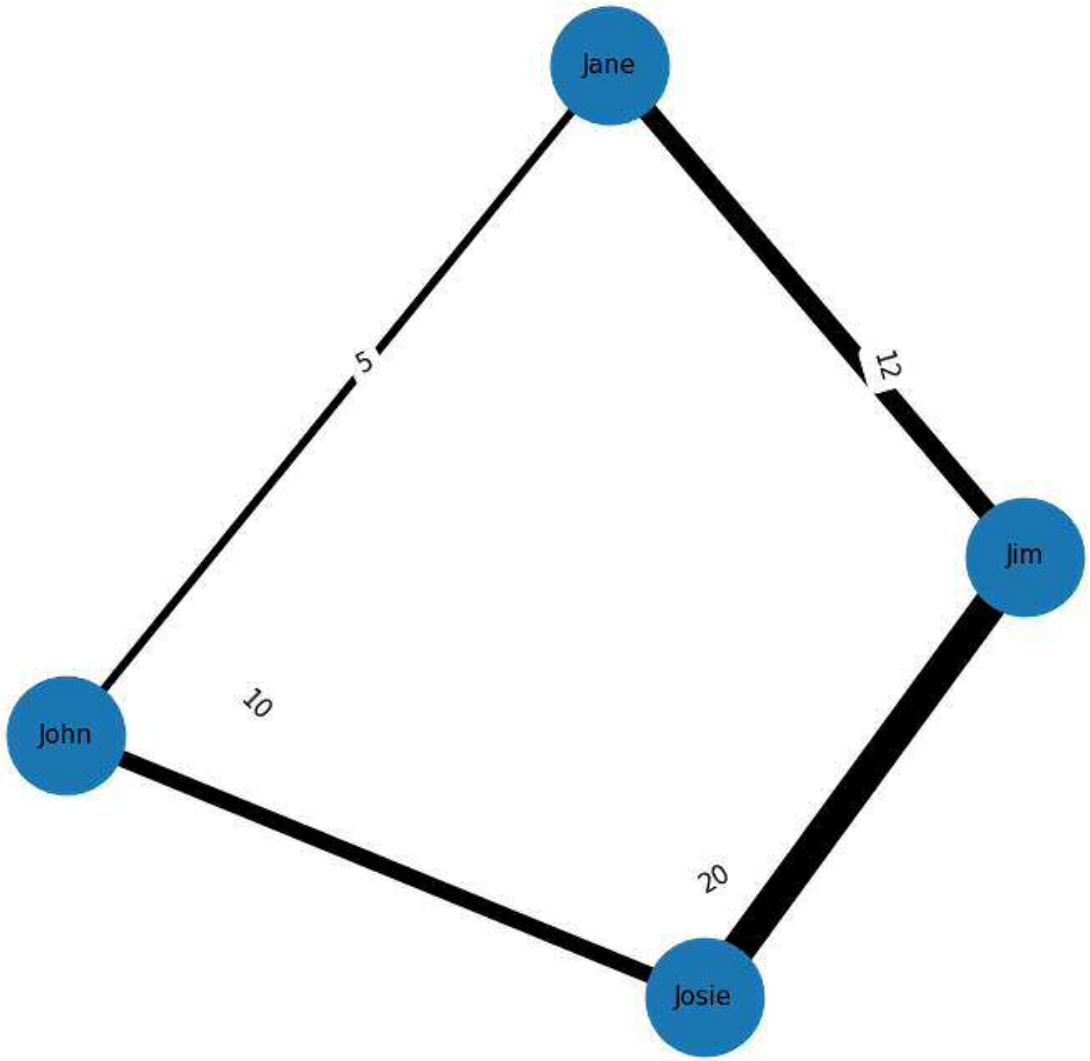
Undirected & Binary Tie



UK Data Service

Connections

Undirected & Valued Tie

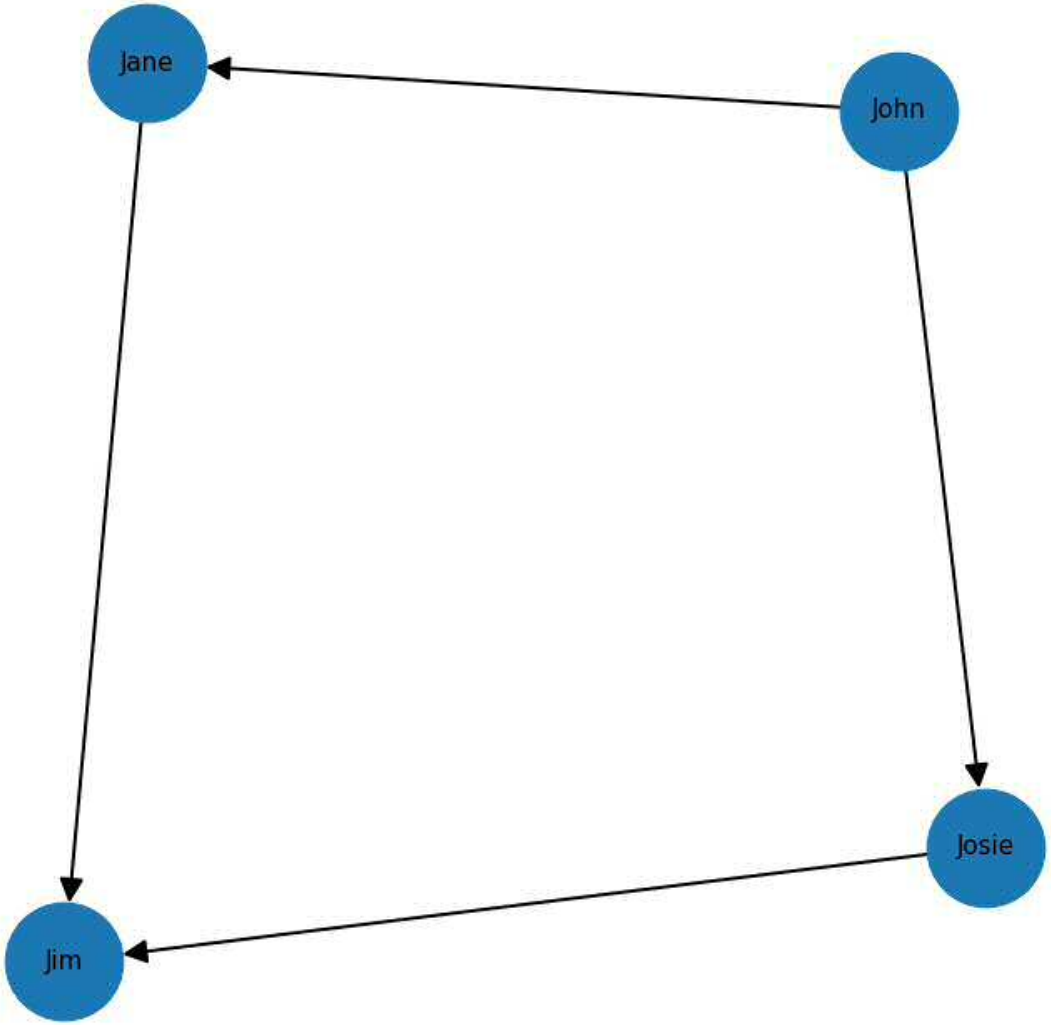


UK Data Service

Connections

Directed &

Binary Tie

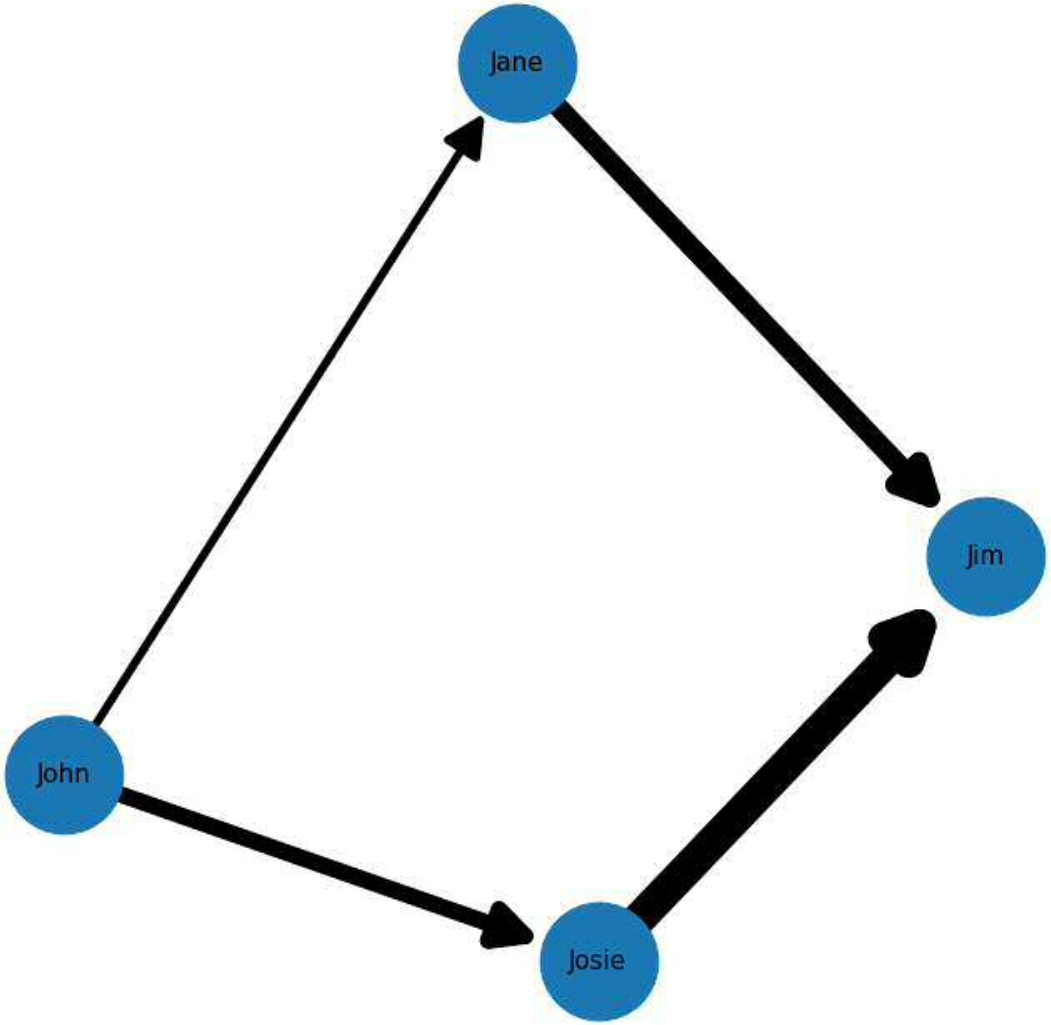


UK Data Service

Connections

Directed &

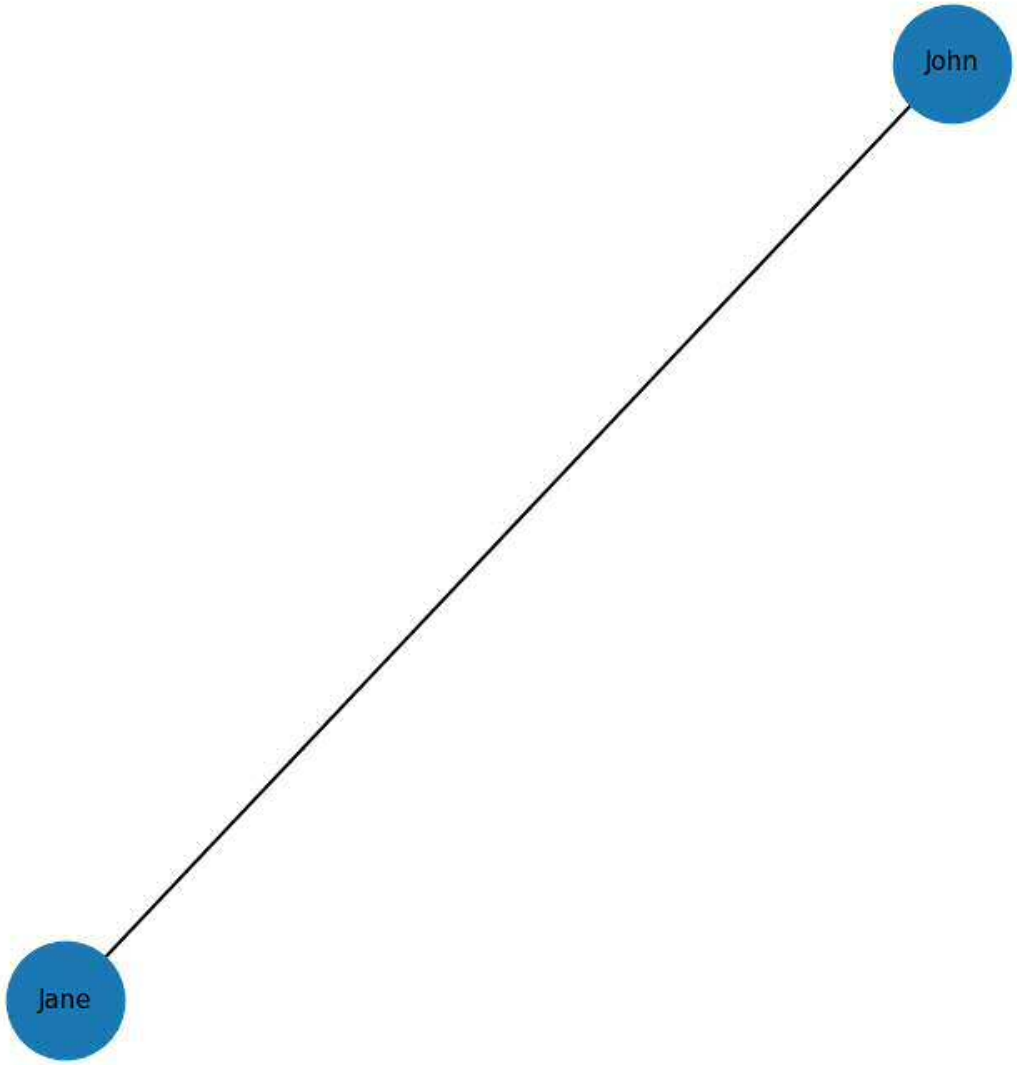
Valued Tie



UK Data Service

Connections

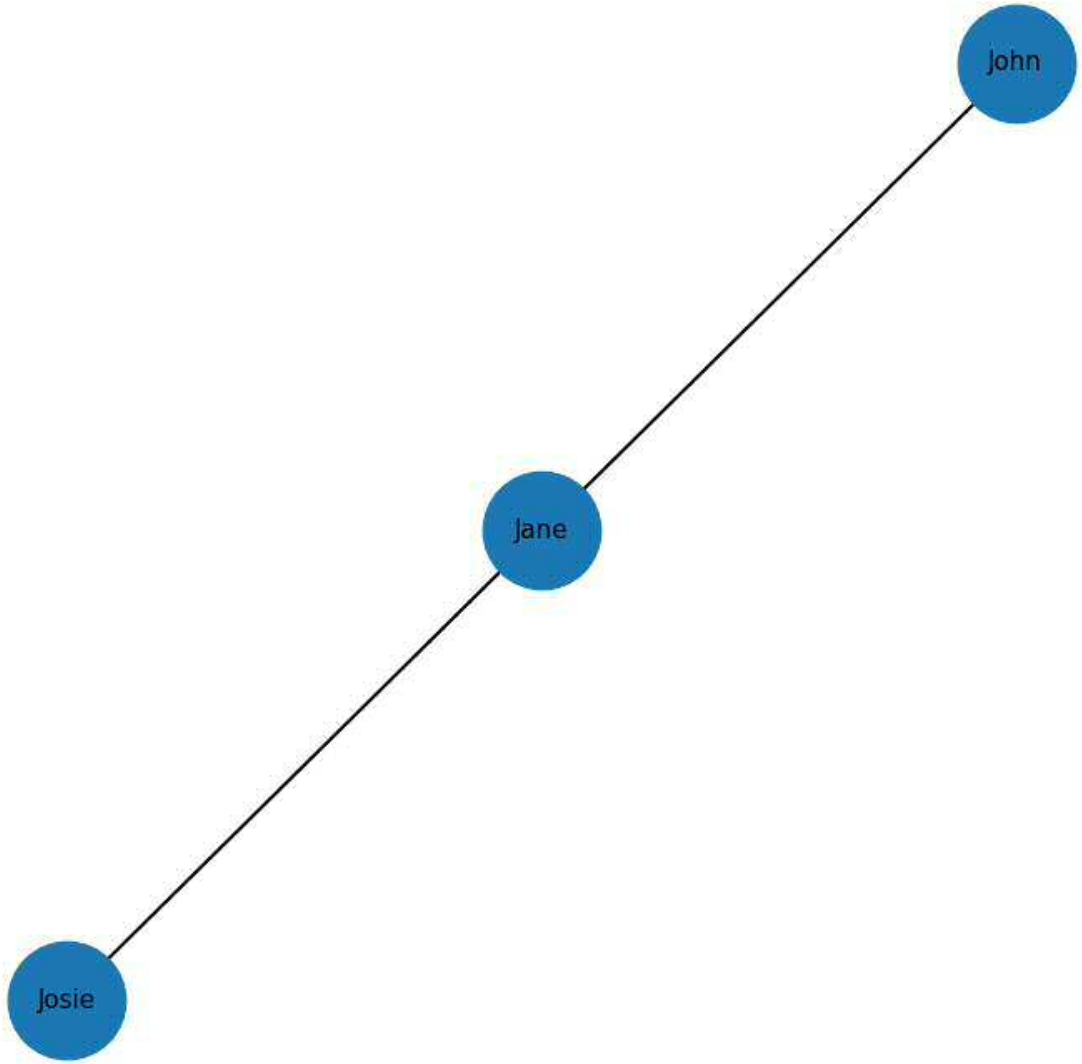
Direct Tie



UK Data Service

Connections

Indirect Tie



UK Data Service

Representing networks

UK Data Service

Representing networks

Networks can be represented using three formats:

1. Matrices
2. Edgelist
3. Graphs

Matrices

A matrix (X) is an arrangement of elements into rows (i) and columns (j).

Social networks can be represented as matrices also:

- Every row is a node
- Every column is a node
- And every value indicates whether a tie exists between two nodes

Let's take a small but real social network as an example.

Matrices

Directed & Binary Network

Adjacency matrix

	Wife	Aunt	Cousin	Gran	Sister-in-law
Wife	--	0	1	0	1
Aunt	1	--	0	1	0
Cousin	1	0	--	1	0
Gran	1	0	1	--	0
Sister-in-law	1	0	0	0	--

UK Data Service

Matrices

Directed & Valued Network

Adjacency matrix

	Wife	Aunt	Cousin	Gran	Sister-in-law
Wife	--	0	2	0	1
Aunt	3	--	0	2	0
Cousin	4	0	--	3	0
Gran	1	0	3	--	0
Sister-in-law	2	0	0	0	--

Edgelist

An edgelist is simply a list of the ties in a network, with the ties represented as pairs of nodes.

source	target	weight
Wife	Cousin	2
Wife	Sister-in-law	1
Aunt	Wife	3
Aunt	Gran	2
Cousin	Wife	4
Cousin	Gran	3
Gran	Wife	1
Gran	Cousin	3
Sister-in-law	Wife	2

Graphs

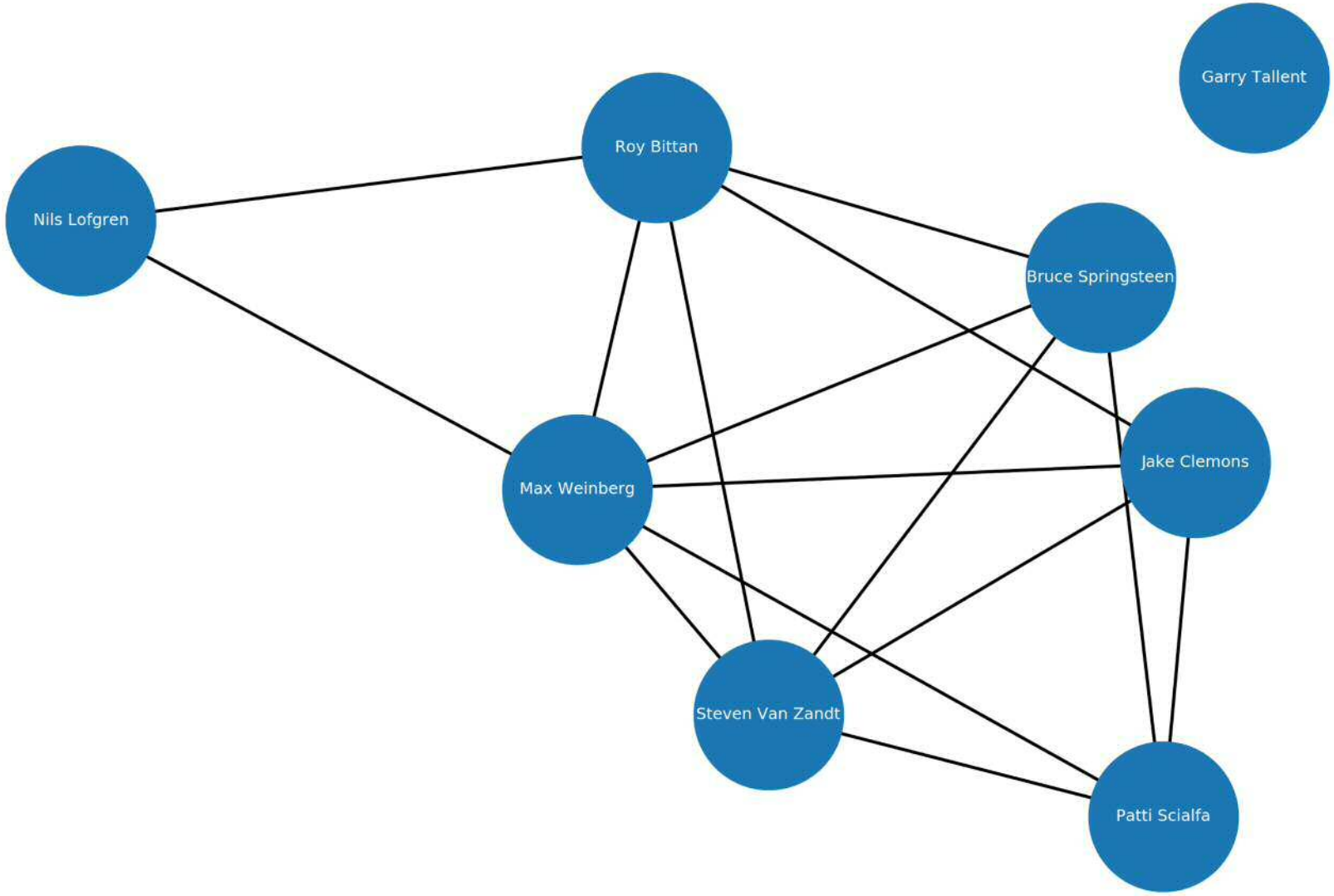
Visual representation of a social network is known as a **graph** or **sociogram**.

A graph is a set of lines connecting points.

Hanneman & Riddle (2005):

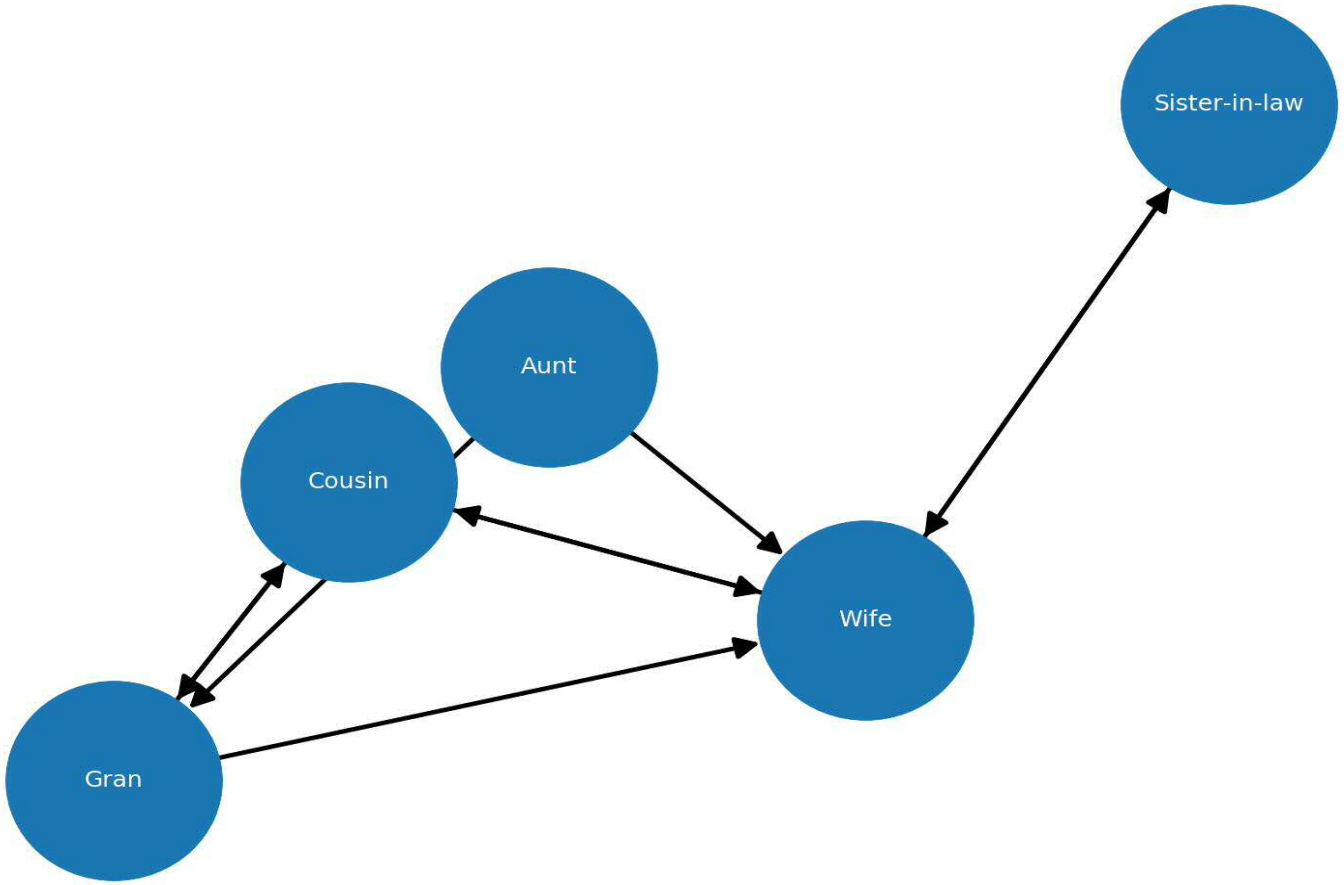
- Nodes are represented as circles
- Ties are represented as lines (with arrow heads if the tie is directed)
- Colours, shapes and sizes can be used to differentiate nodes by their attributes or network characteristics.
- Colours, shapes and sizes can also be used to differentiate relations by their type or amount.

Graphs



UK Data Service

Graphs



UK Data Service

A simple analysis

A simple analysis

Research question: What degree of board interlock occurs in the UK charity sector? Board interlock is a measure of the degree to which organisations are connected through shared board members.

Nodes and connections: Registered charities and whether they have trustees in common. That is, two charities are connected if they have at least one individual who acts as a trustee of both organisations.

Data set: Current trustees of charities headquartered in Manchester.

Analysis: Interested in analysing the size of the network, how cohesive it is, and which charities are the most connected.

Questions

Dr Diarmuid McDonnell

diarmuid.mcdonnell@manchester.ac.uk

@DiarmuidMc

Further resources and help

Repository: https://github.com/UKDataServiceOpen/new-forms-of-data

Youtube: https://www.youtube.com/user/UKDATASERVICE

Help: ukdataservice.ac.uk/help/

Subscribe to UK Data Service news at https://www.jiscmail.ac.uk

@UKDataService

UKDataService

UK Data Service