

345 Ludic Computing

Tutorial 7

# Social Networks

Simon Colton & Alison Pease

Computational Creativity Group

Department of Computing

Imperial College London

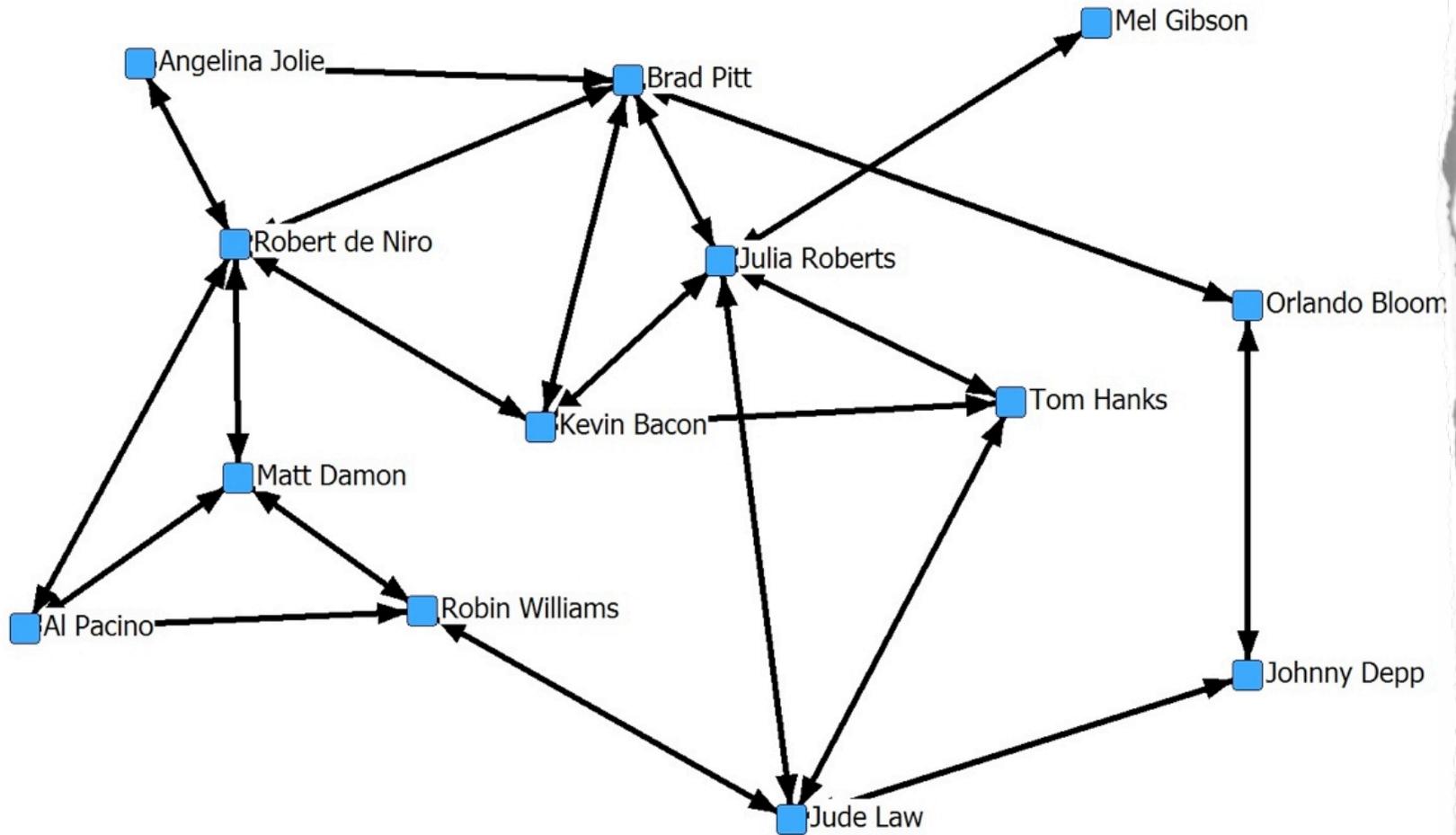
[ccg.doc.ic.ac.uk](http://ccg.doc.ic.ac.uk)

# Question 1

Consider the actors network on the following slide

- a) What is the diameter of the network?
- b) What is the closeness centrality of Julia Roberts?
- c) What is the clustering coefficient of Brad Pitt?

# Question 1



# Question 1

## Some Useful Definitions

- Neighbourhood of  $n_i$  ( $\Gamma_{n_i}$ ): subgraph formed by nodes adjacent to  $n_i$
- Diameter (of a graph): longest shortest path length between any two nodes in  $G$
- Closeness centrality: sum of geodesic distances to all other nodes
- Clustering Coefficient  $C(n_i)$ : measures the connectivity among those  $k_{n_i}$  nodes which are in the neighbourhood of  $n_i$

$$C(n_i) = \frac{2|E(\Gamma_{n_i})|}{k_{n_i}(k_{n_i} - 1)}$$

$|E(\Gamma_{n_i})|$ : number of edges in the neighbourhood of  $n_i$

# Answer 1

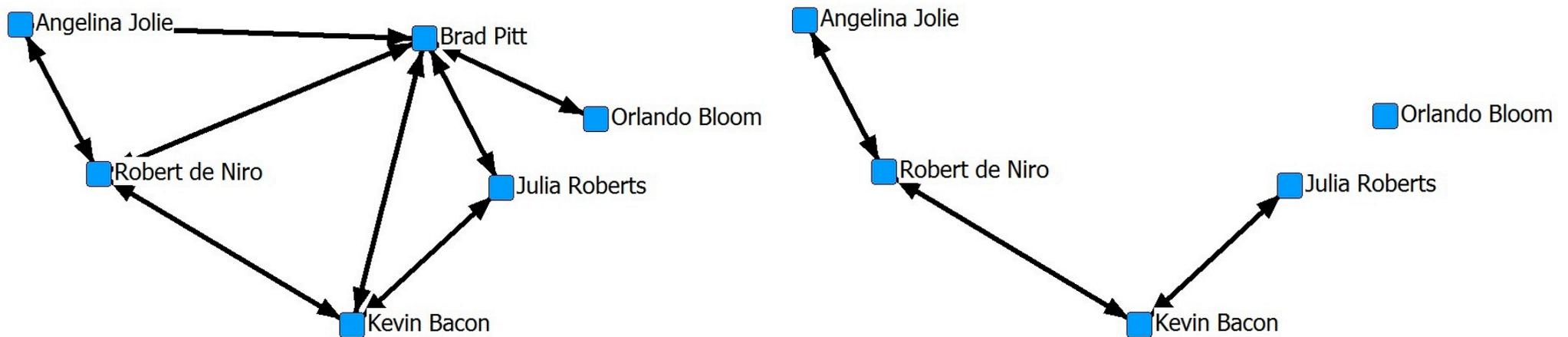
- a) The longest shortest path length is the geodesic distance between Matt Damon and Mel Gibson or Al Pacino and Mel Gibson, i.e. 4
- b) First we need to find the geodesic distance from Julia Roberts to all other nodes:

Johnny Depp: 2	Kevin Bacon: 1	Matt Damon: 3
Robin Williams: 2	Tom Hanks: 1	Mel Gibson: 1
Angelina Jolie: 2	Jude Law: 1	Al Pacino: 3
Brad Pitt: 1	Robert de Niro: 2	Orlando Bloom: 2

Then we sum all the geodesic distances = 21

# Answer 1

c) First we need to extract the neighbourhood of Brad Pitt



Note that the neighbourhood does not include the node itself.

Then we work out the number of edges in the neighbourhood, i.e.  $|E(\Gamma_{n_i})| = 3$  and  $k_{n_i} = 5$ . Thus,  $C(n_i) = \frac{2(3)}{5(5-1)} = 0.3$

# Question 2

Page Rank is a ranking algorithm originally proposed by Sergey Brin and Lawrence Page which they used to return search results in Google.

The page rank of a page  $A$  is calculated as:

$$PR(A) = (1 - d) + d(PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

where:

$T1 \dots Tn$ : pages pointing to  $A$

$C(A)$ : number of links going out of page  $A$

$d$ : damping factor -  $[0, 1]$

What do you consider are the key characteristics of Page Rank? (which, by the way, might help explain Google's advantage over similar search engines.)

# Answer 2

Key characteristics of PageRank (among others):

- It not only returns relevant search results but more importantly prioritises them according to their links from important pages.
- Dividing by the number of links going out of a page prevents hub pages from passing on disproportionate rankings.
- PageRank was conceived as the probability that a person who starts on a random page, only surfs forward and if bored jumps to another random page. Thus, the free parameter  $d$  can be used to account for the fact (i.e. probability) that the person might stop surfing at any step.