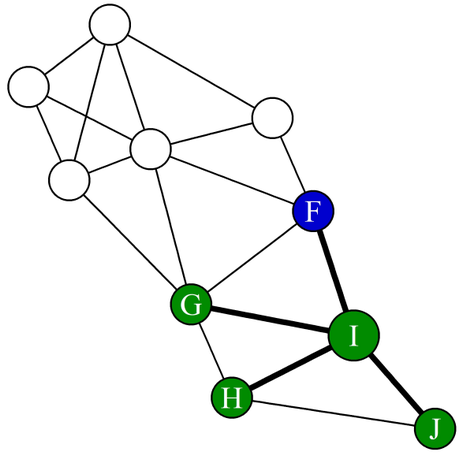
Basic Network features

PREDICTIVE ANALYTICS USING NETWORKED DATA IN R

Bart Baesens, Ph.D.

Professor of Data Science, KU Leuven and University of Southampton

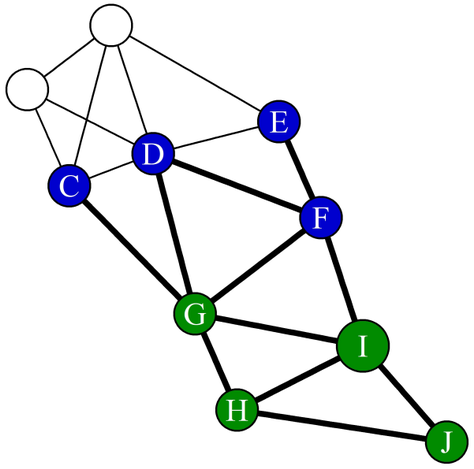
Neighborhood features



- First order degree
  - Number of connected nodes

```
degree(g)
```

```
A B C D E F G H I J
4 3 4 6 3 4 5 3 4 2
```

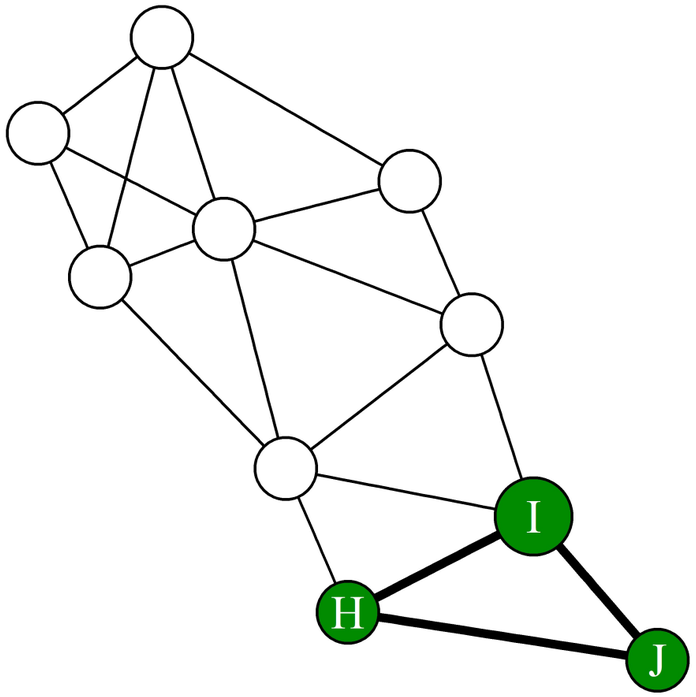


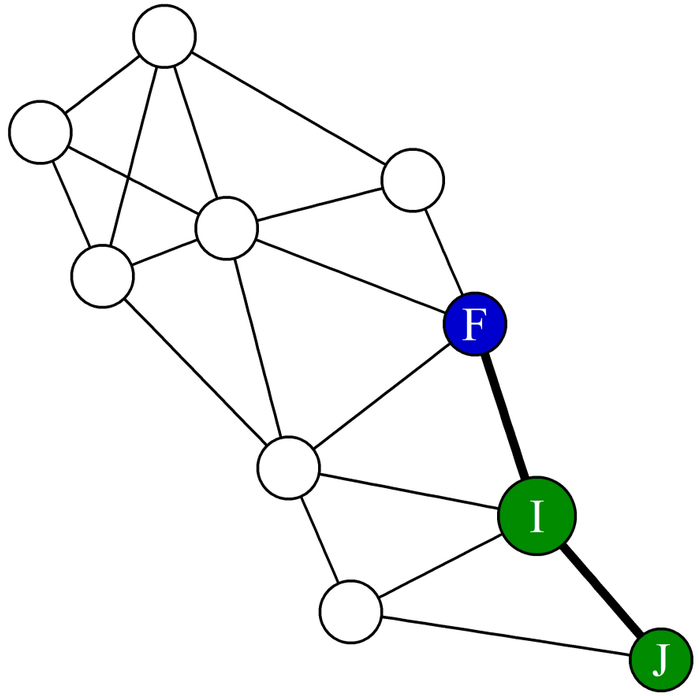
- Second order degree
  - Number of connected nodes that are two or less edges away

```
neighborhood.size(g, order = 2)
```

```
7  7  9  9  8 10 10  7  8  5
```

Neighborhood features - triangles



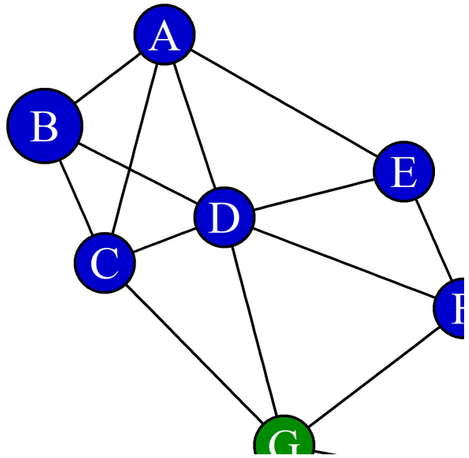



```
count_triangles(g)
```

```
4 3 4 7 2 3 4 2 3 1
```

Centrality features

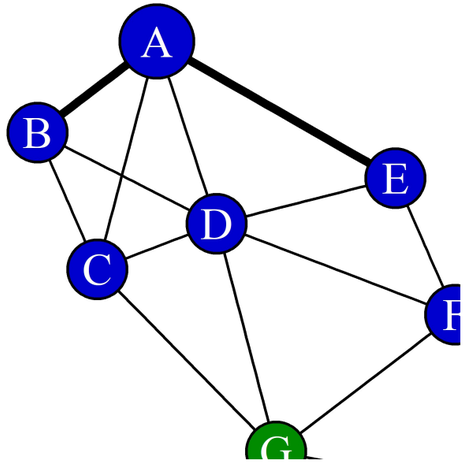
- Betweenness




```
betweenness(g)
```

```
   A    B    C    D    E    F     G    H    I    J
1.00 0.00 3.32 8.10 0.92 5.37 11.47 2.07 5.77 0.00
```

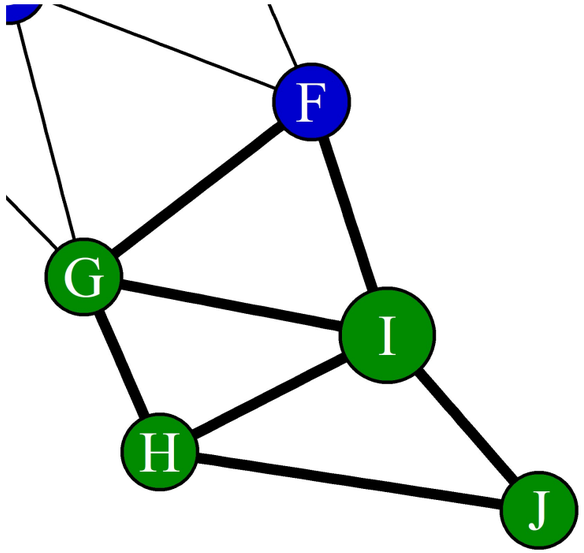
- Closeness

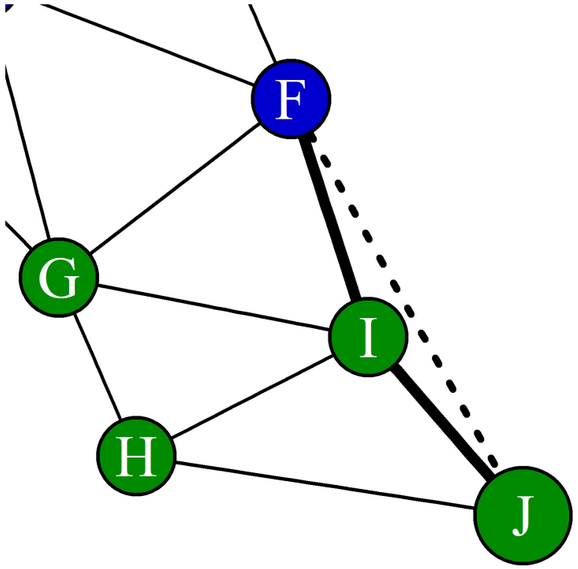



```
closeness(g)
```

```
   A    B    C    D    E    F    G    H    I    J
0.06 0.05 0.07 0.08 0.06 0.07 0.08 0.06 0.06 0.04
```

Transitivity






```
transitivity(g,type = 'local')
```

```
0.67 1.00 0.67 0.47 0.67 0.50 0.40 0.67 0.50 1.00
```

Let's practice!

PREDICTIVE ANALYTICS USING NETWORKED DATA IN R

Link Based Features

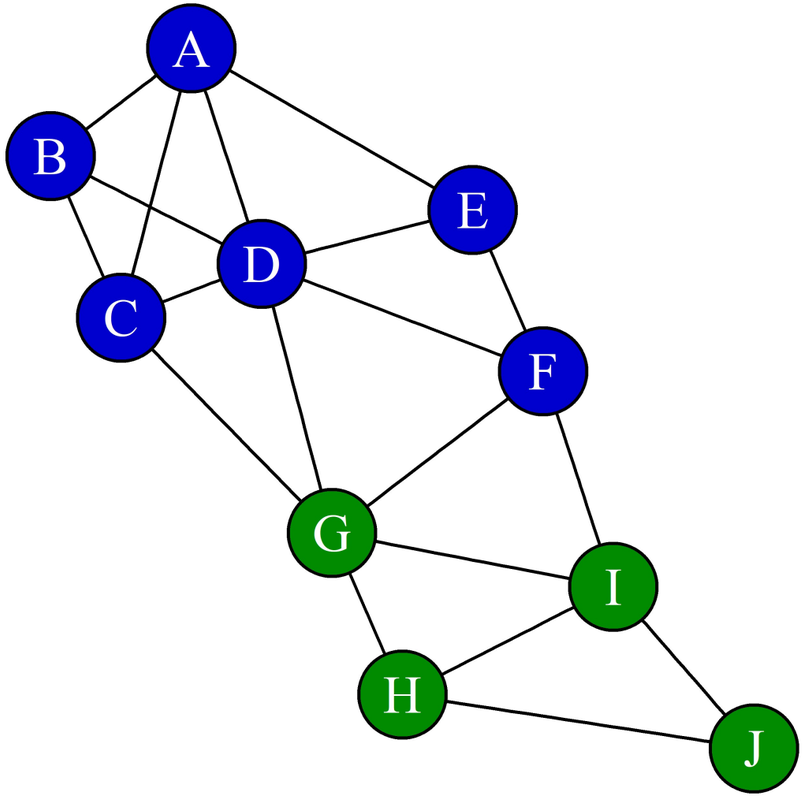
PREDICTIVE ANALYTICS USING NETWORKED DATA IN R

María Óskarsdóttir, Ph.D.
Post-doctoral researcher

datacamp

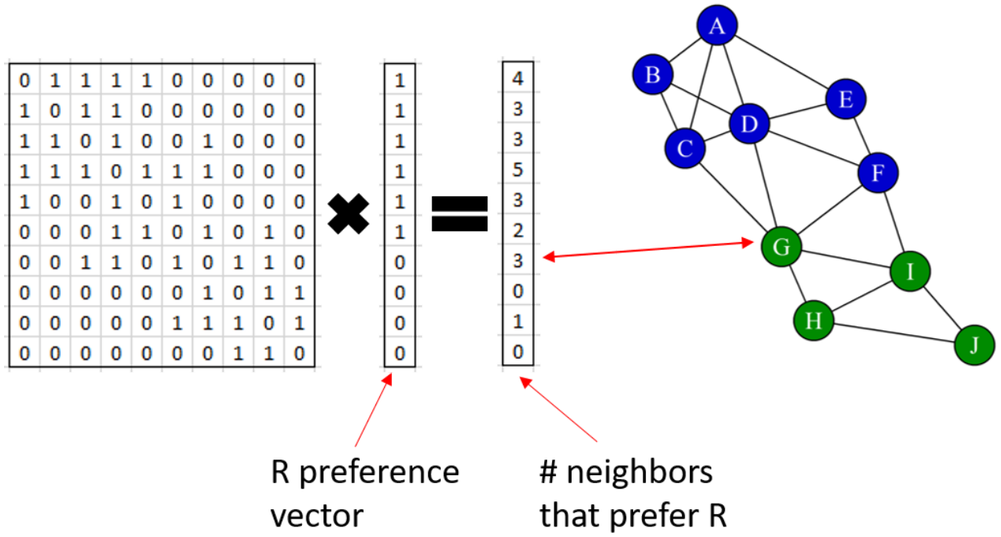
Adjacency matrices

	A	B	C	D	E	F	G	H	I	J
A	0	1	1	1	1	0	0	0	0	0
B	1	0	1	1	0	0	0	0	0	0
C	1	1	0	1	0	0	1	0	0	0
D	1	1	1	0	1	1	1	0	0	0
E	1	0	0	1	0	1	0	0	0	0
F	0	0	0	1	1	0	1	0	1	0
G	0	0	1	1	0	1	0	1	1	0
H	0	0	0	0	0	0	1	0	1	1
I	0	0	0	0	0	1	1	1	0	1
J	0	0	0	0	0	0	0	1	1	0




```
A <- get.adjacency(g)
```

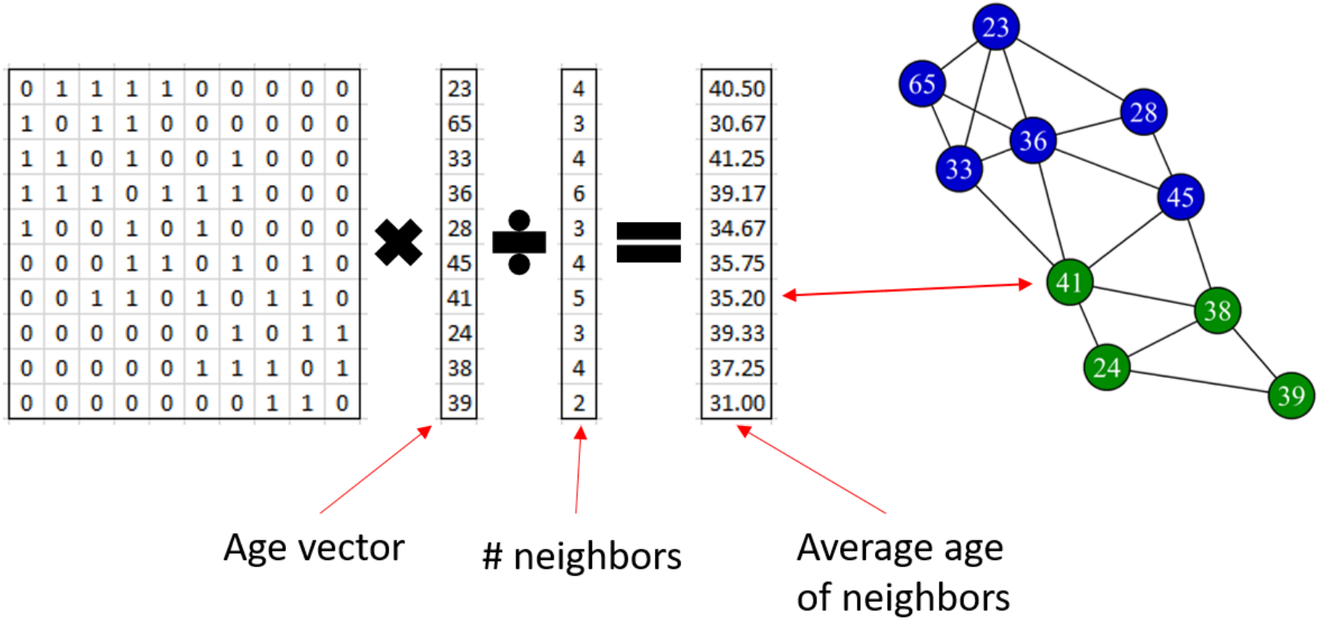
Link based features




```
preference <- c(1,1,1,1,1,1,0,0,0,0)
rNeighbors <- A %*% preference
as.vector(rNeighbors)
```

```
4 3 3 5 3 2 3 0 1 0
```

Neighborhood features




```
age <- c(23,65,33,36,28,45,41,24,38,39)
degree <- degree(g)
averageAge <- A %*% age / degree
```

Let's practice!

PREDICTIVE ANALYTICS USING NETWORKED DATA IN R

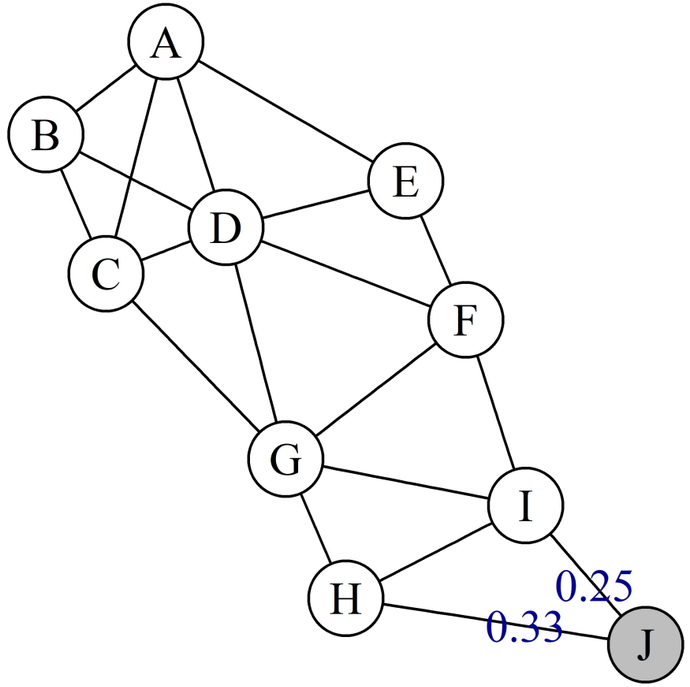
PageRank

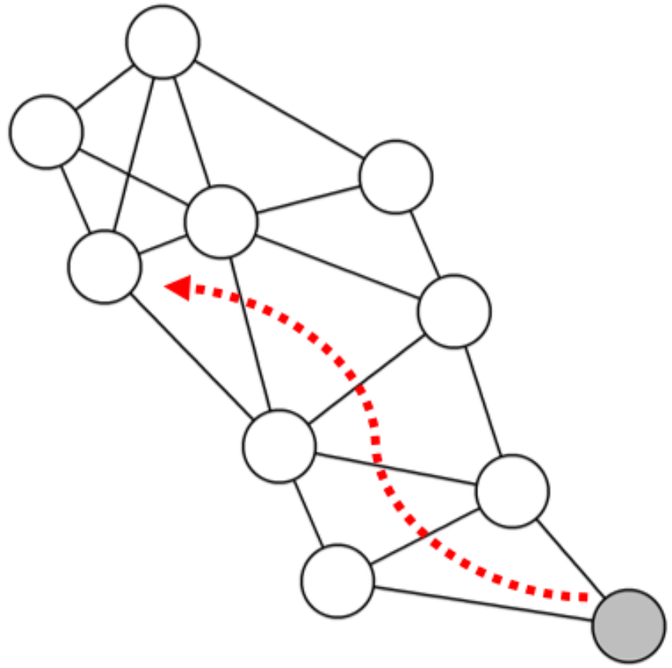
PREDICTIVE ANALYTICS USING NETWORKED DATA IN R

María Óskarsdóttir, Ph.D.
Post-doctoral researcher

datacamp

The PageRank algorithm





$$\text{PageRank}_J = \alpha \cdot (\tfrac{1}{3} \cdot \text{PageRank}_H + \tfrac{1}{4} \cdot \text{PageRank}_I) + (1 - \alpha) \cdot e_J$$

The PageRank algorithm

$$\vec{PR} = \alpha \cdot A \cdot \vec{PR} + (1 - \alpha) \cdot \vec{e}$$

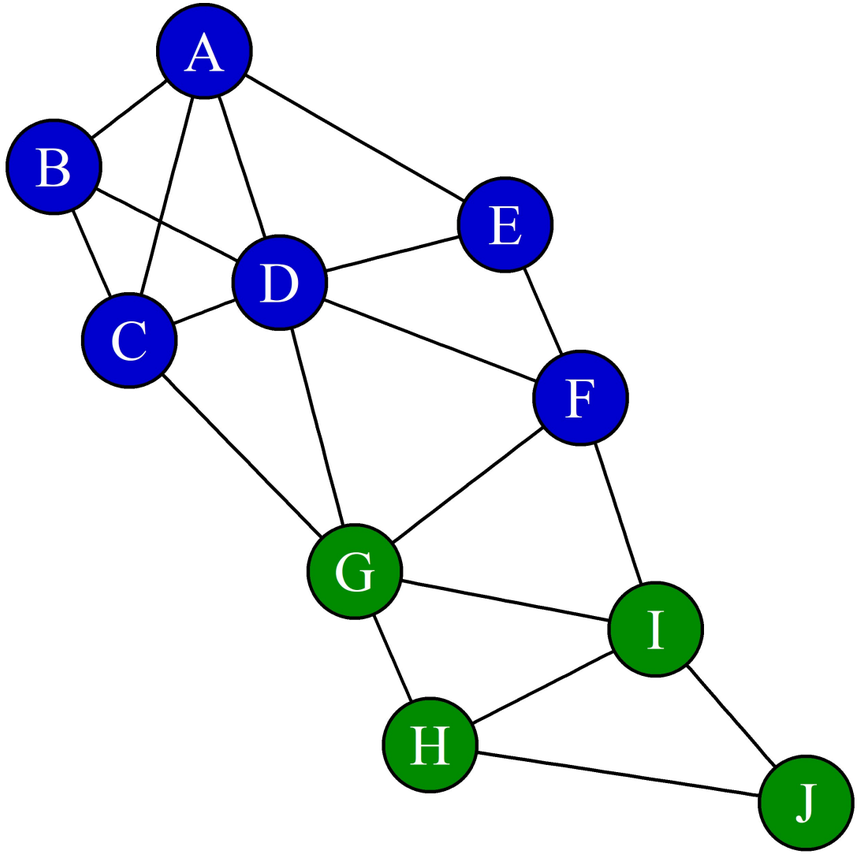
```
page.rank(g)
```

```
$vector
         A          B          C          D          E          F          G
0.10238312 0.07917232 0.10164910 0.14693274 0.07953551 0.10335821 0.12732387
         H          I          J
0.08675903 0.10994175 0.06294435

$value
[1] 1

$options
NULL
```

Personalized PageRank




```
page.rank(g,
 personalized = c(1,0,0,0,0,0,0,0,0,0))
```

```
$vector
         A          B          C          D          E
0.25528911 0.10363533 0.12156935 0.16625582 0.09366836
         F          G          H          I          J
0.07466596 0.08473039 0.03285162 0.04785657 0.01947748

$value
[1] 1

$options
NULL
```

Let's practice!

PREDICTIVE ANALYTICS USING NETWORKED DATA IN R