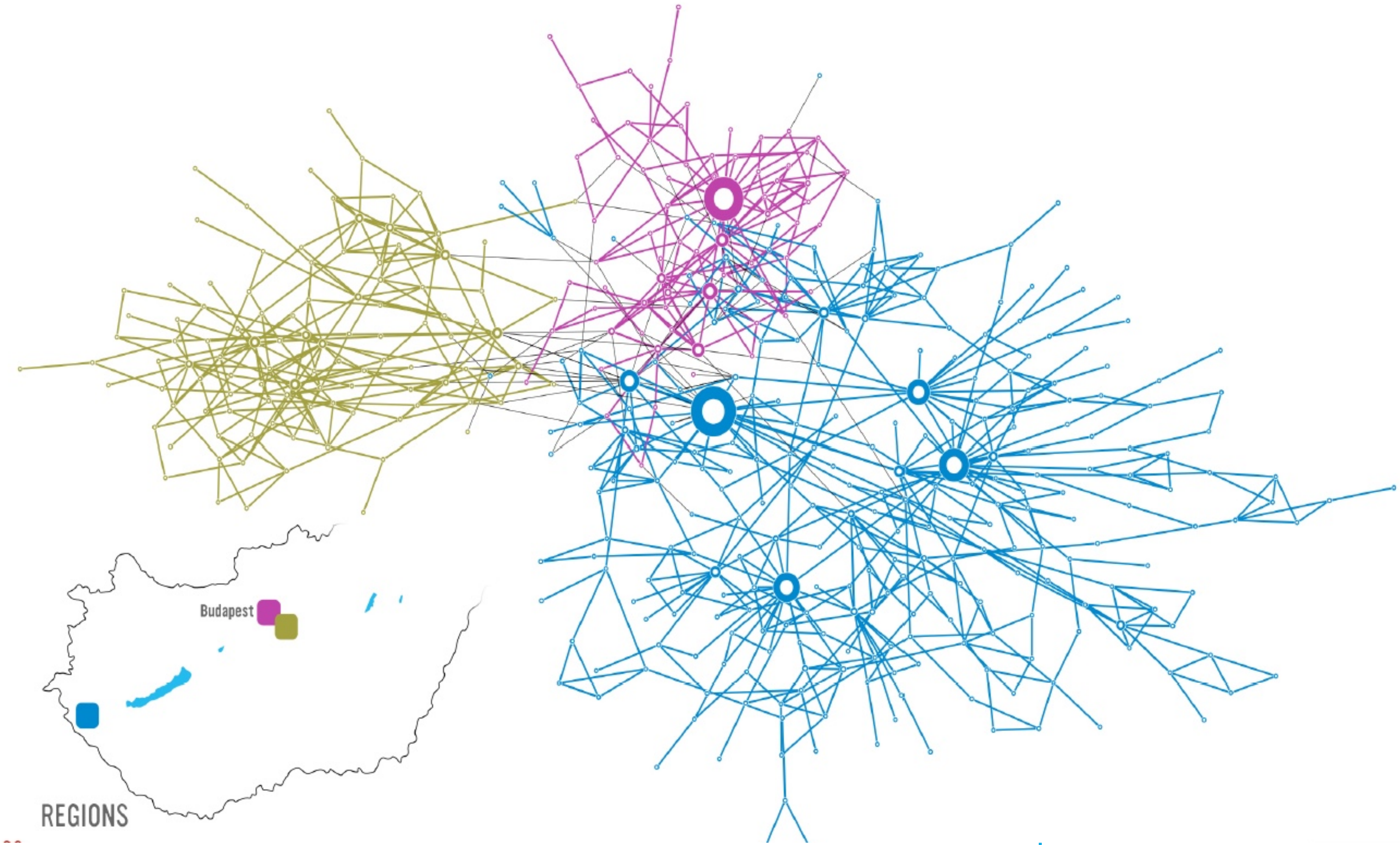


NETWORK SCIENCE

Network Science
Summer Research Institute 2019

IMAGE FROM ALBERT-LASZLO BARABASI'S *NETWORK SCIENCE.* (2016) CAMBRIDGE UNIVERSITY PRESS.

INTRODUCTIONS

THE WEEK

Monday

- Introduction to Network Science
- The Princess Bride
- Crash course R programming

Tuesday

- Crash course in sentiment analysis and regular expressions
- Crash course in network measures and community detection
- Choose your movie and form research questions and hypotheses

Wednesday

- Present movie choices
- Data extraction

Thursday

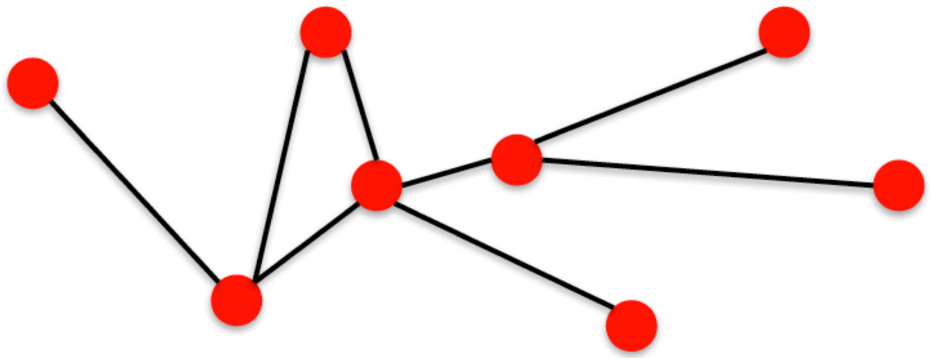
- Data extraction
- Network and sentiment analysis

Friday

- Analysis
- Presentation prep.
- Presentation

WHAT IS A NETWORK?

A set of points joined in pairs by lines.



Network Science	Graph Theory
Network	Graph
Node	Vertex
Link	Edge
Often refers to real systems.	Mathematical representation of a network

MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

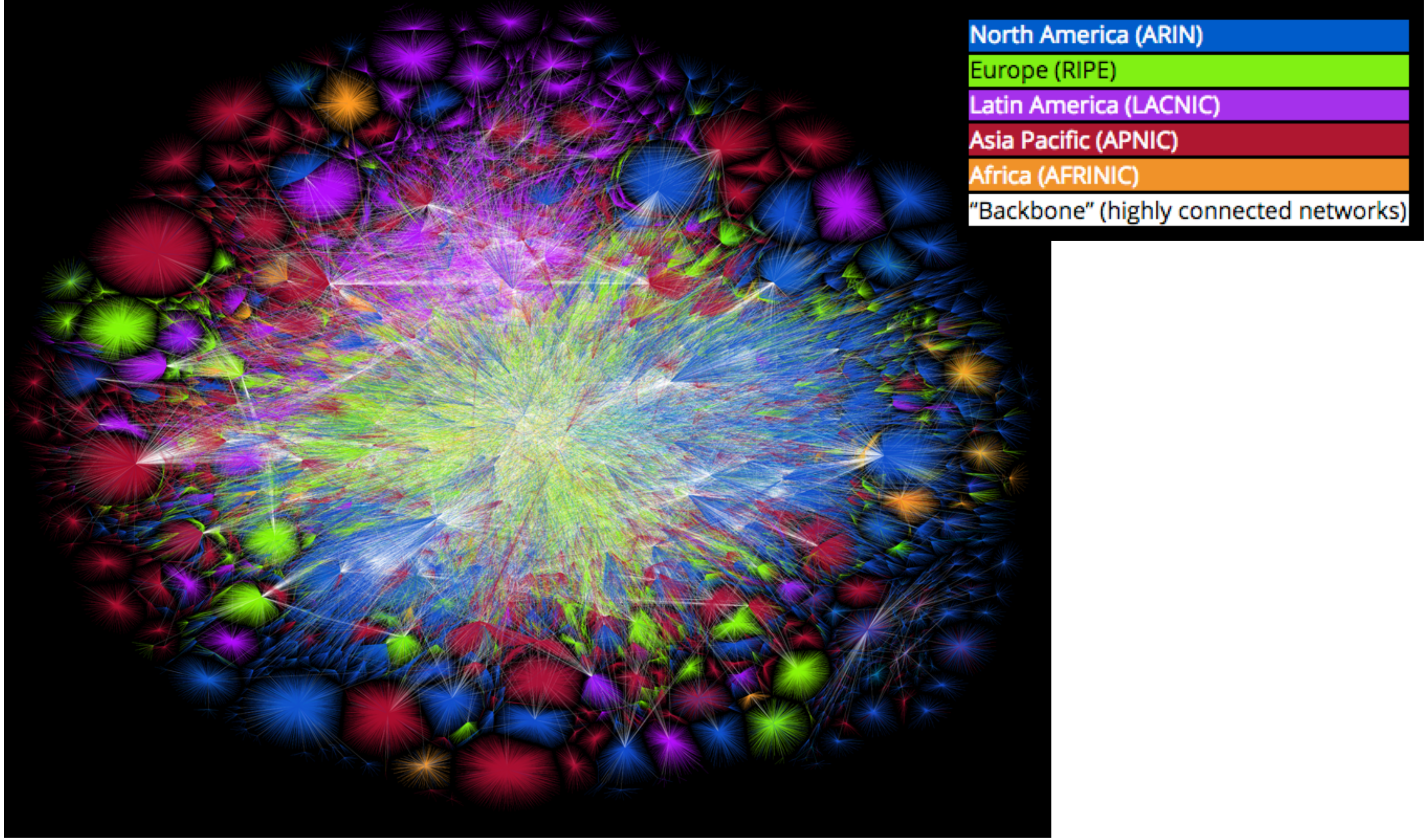
M.E.J. NEWMAN. *NETWORKS AN INTRODUCTION.* (2010) OXFORD UNIVERSITY PRESS

WHAT IS NETWORK SCIENCE?

The study of complex systems through a network which encodes the interactions between components.

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	$\langle k \rangle$
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

TABLE FROM ALBERT-LASZLO BARABASI'S *NETWORK SCIENCE.* (2016) CAMBRIDGE UNIVERSITY PRESS.



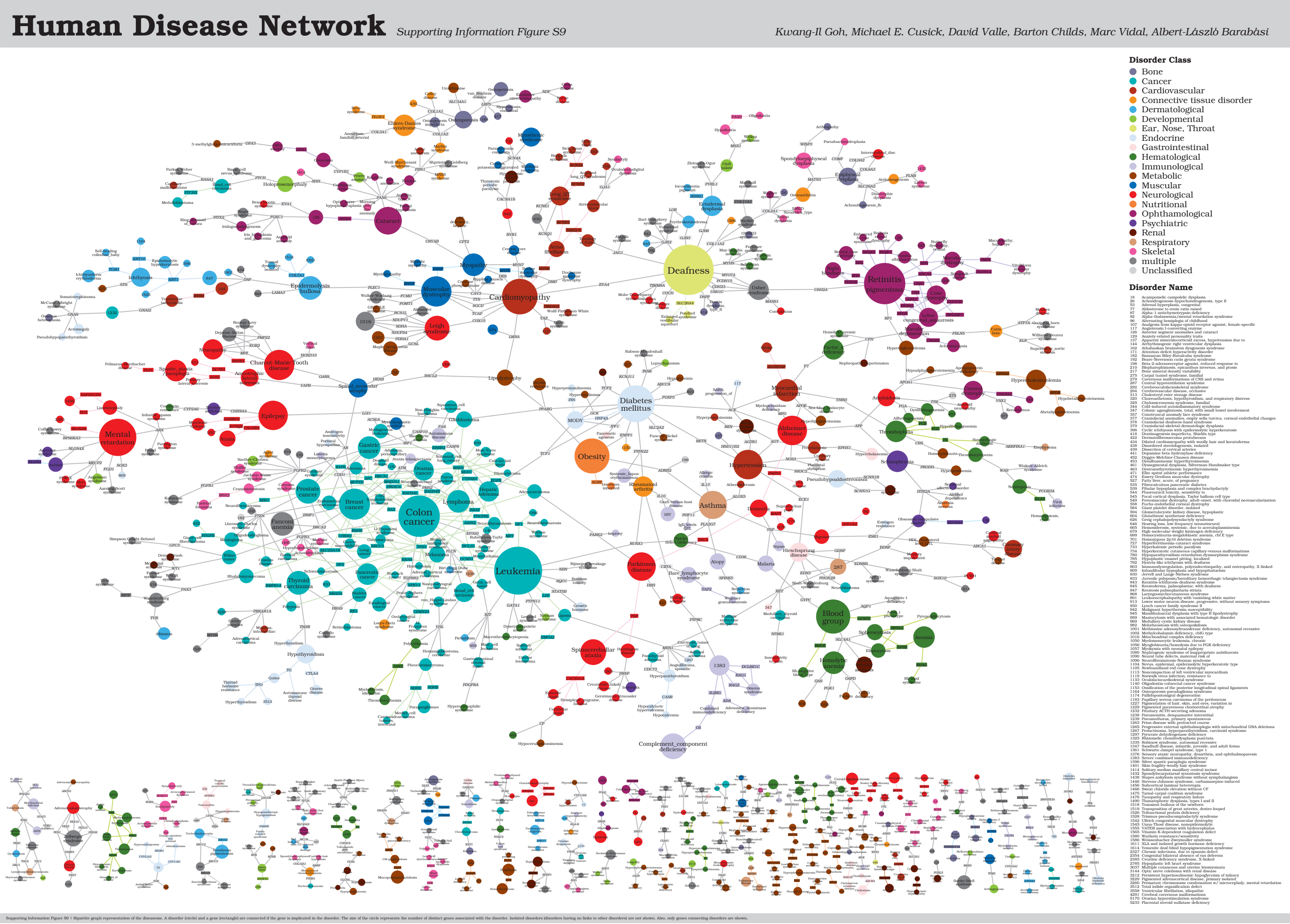
THE INTERNET

Visualization of the routing paths of the Internet. Barrett Lyon/The Opte Project, July 11, 2015

WWW.OPTE.ORG

FACEBOOK

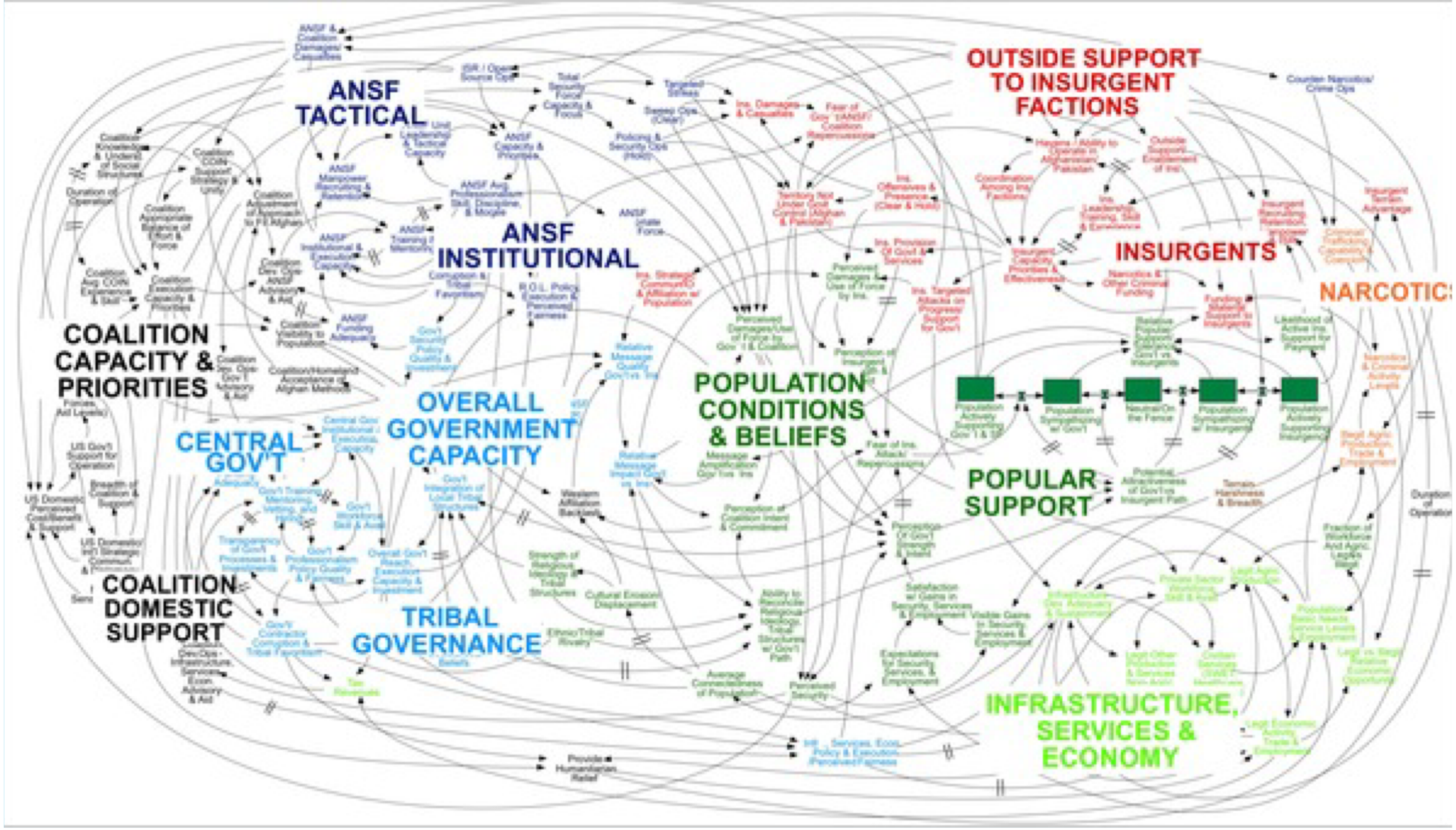
HTTPS://MEDIUM.COM/@JOHNROBB/FACEBOOK-THE-COMPLETE-SOCIAL-GRAPH-B58157EE6594



HUMAN DISEASE NETWORK

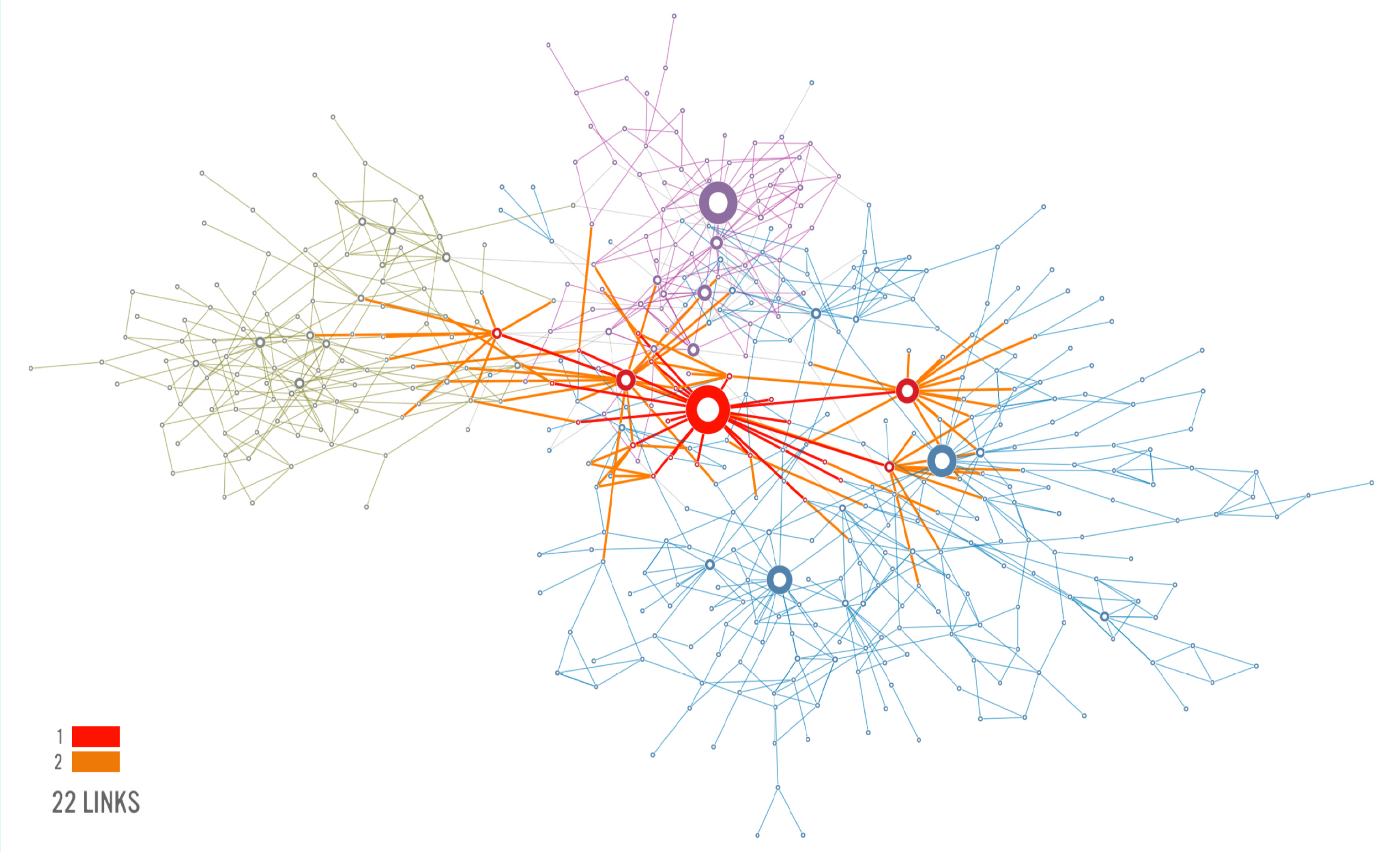
Diseases are connected if they have a common genetic origin.

IMAGE FROM ALBERT-LASZLO BARABASI'S *NETWORK SCIENCE*. (2016) CAMBRIDGE UNIVERSITY PRESS.



MILITARY ENGAGEMENT

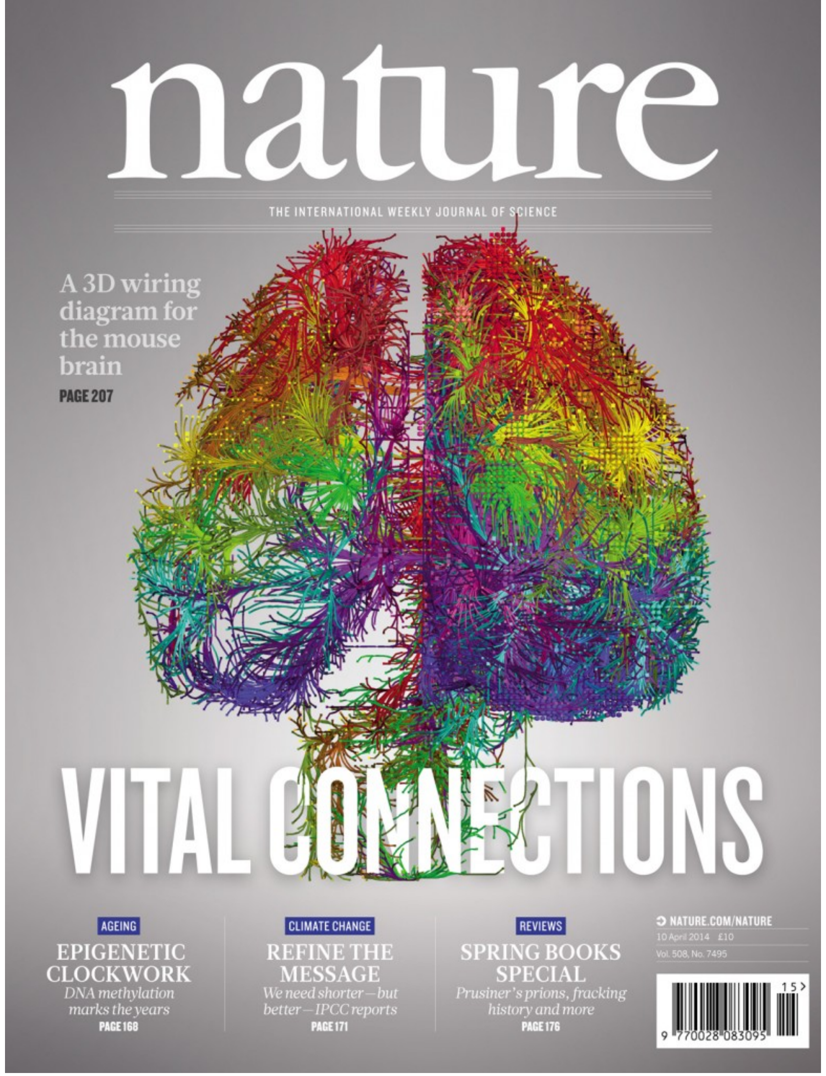
Designed during the Afghan war, 2012



COMPANY NETWORK

People are connected if one nominated the other as a source of information about organizational and professional issues.

ALBERT-LASZLO BARABASI'S *NETWORK SCIENCE.* (2016) CAMBRIDGE UNIVERSITY PRESS.



BRAIN NETWORK

April 10, 2014 Issue of *Nature*, neuronal connectivity in a mouse.

SAMPLE NETWORK SCIENCE APPLICATIONS

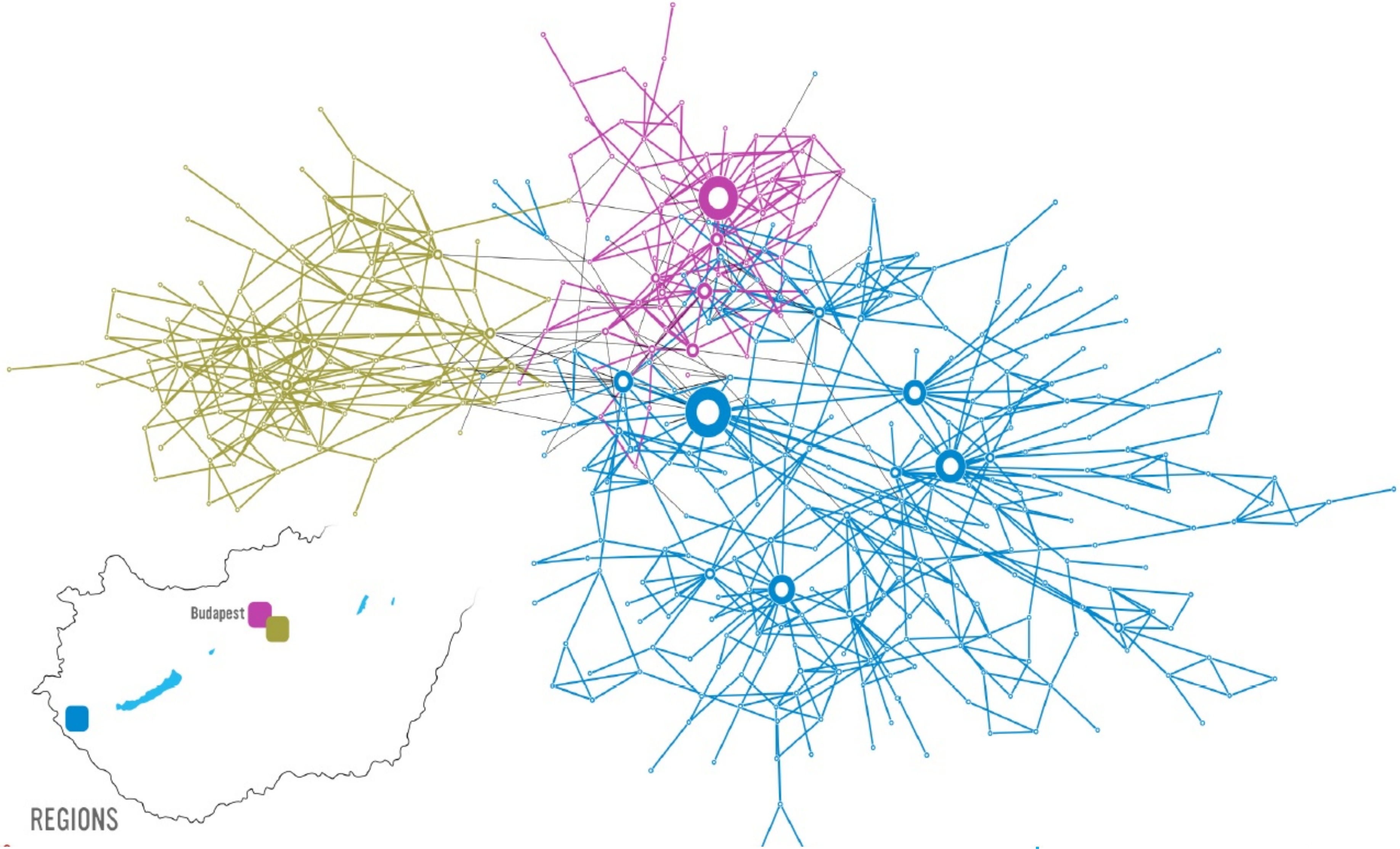
Network	Application
WWW	What web pages are most related to a search term?
Power Grid	What areas are vulnerable to power failures?
Protein Interactions	How do protein interactions impact human health?
Social / Company Networks	How does information spread?

WHAT MAKES A CHARACTER IMPORTANT?

PRINCESS BRIDE

Make a network from the data you collected.

Who is the most important character? Why?



NETWORK MEASURES

Network Science
Summer Research Institute 2019

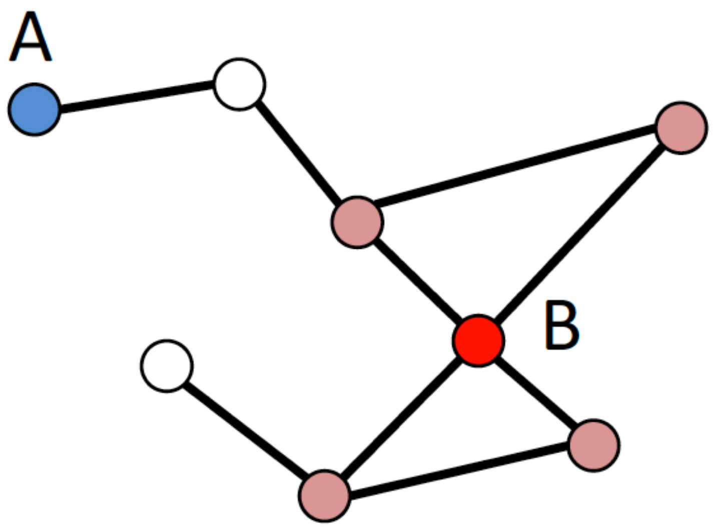
IMAGE FROM ALBERT-LASZLO BARABASI'S *NETWORK SCIENCE*. (2016) CAMBRIDGE UNIVERSITY PRESS.

NODE DEGREE

The number of links connected to the node.

$k_A = 1$

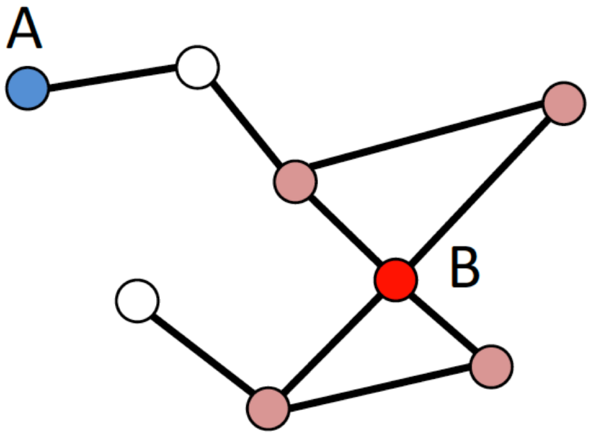
$k_B = 4$



AVERAGE DEGREE

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N}$$

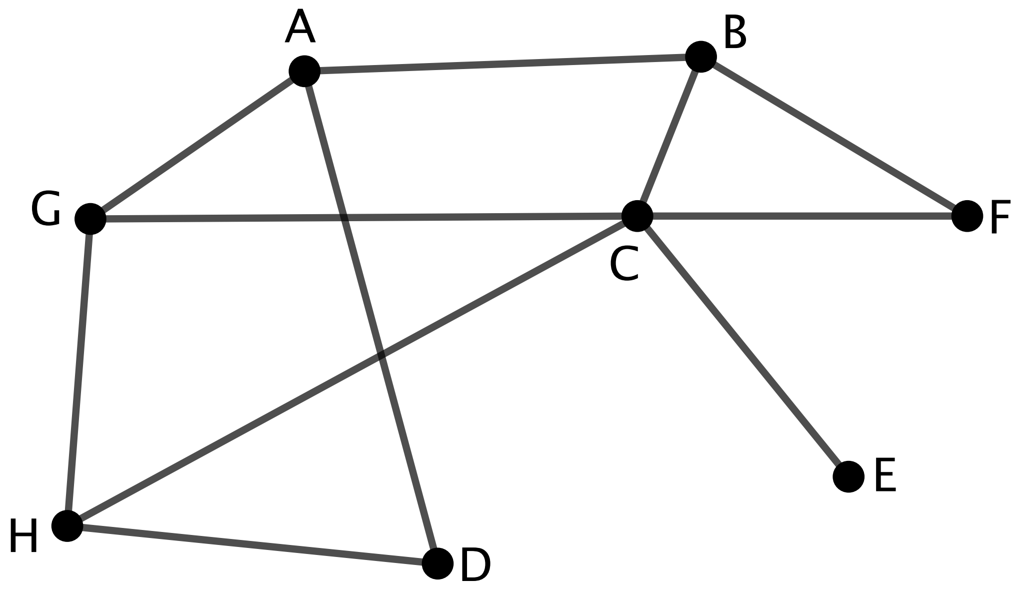
N is the number of nodes
L is the number of links



$$\langle k \rangle = \frac{2 \cdot 9}{8} = \frac{9}{4}$$

AVERAGE DEGREE

Find the degree of each node in the network and the average degree of the network.



AVERAGE DEGREE

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

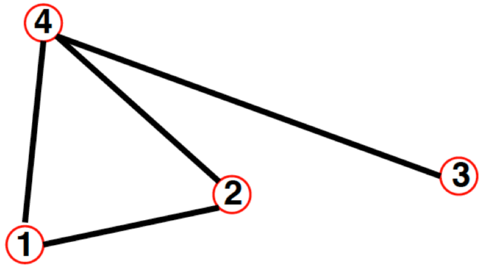
MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

ADJACENCY MATRIX

For a network with *n* nodes, we form an *nxn* matrix, *A*, such that

- $A_{ij} = 1$ if there is a link between node i and j
- $A_{ij} = 0$ if there is no link between node i and j

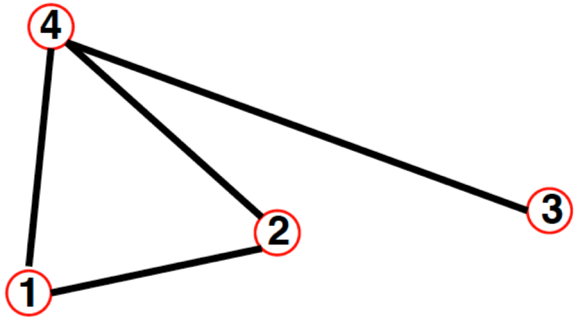
Example



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

ADJACENCY MATRIX AND DEGREES



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ij} = A_{ji}$$
$$A_{ii} = 0$$

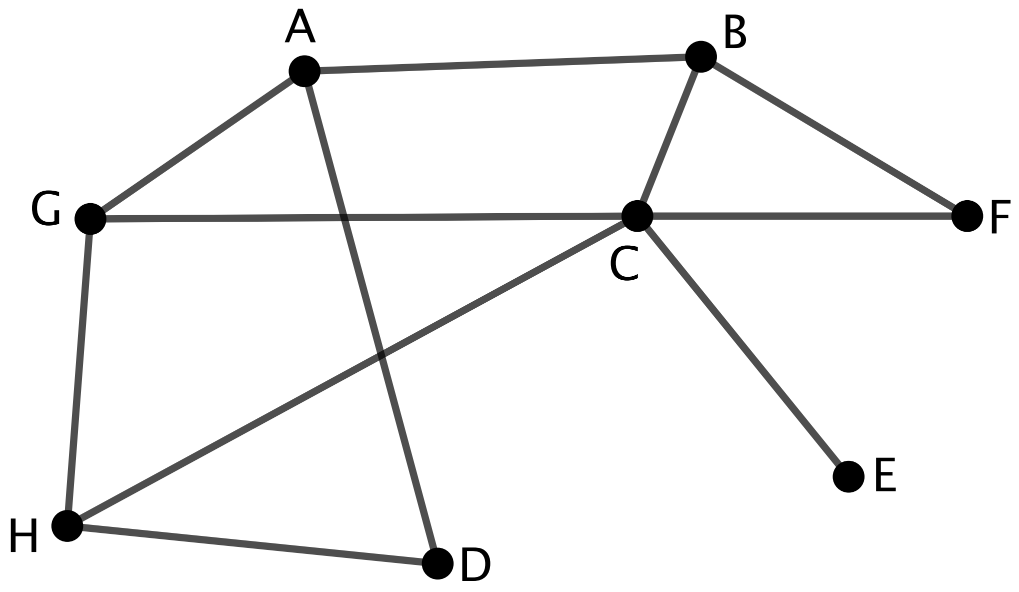
$$\boxed{k_i = \sum_{j=1}^{N} A_{ij}}$$

$$\boxed{k_j = \sum_{i=1}^{N} A_{ij}}$$

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i = \frac{1}{2}\sum_{ij}^{N} A_{ij}$$

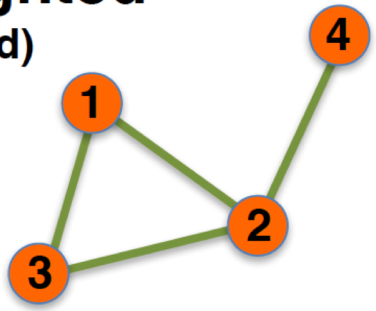
ADJACENCY MATRIX

Find the adjacency matrix for the network.



WEIGHTED GRAPHS

Unweighted
(undirected)

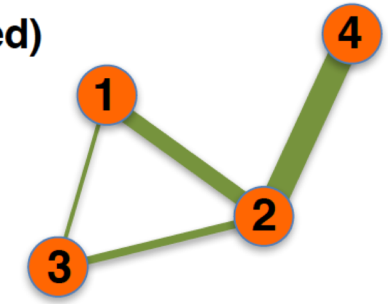


$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$$

Weighted
(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad <k> = \frac{2L}{N}$$

MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

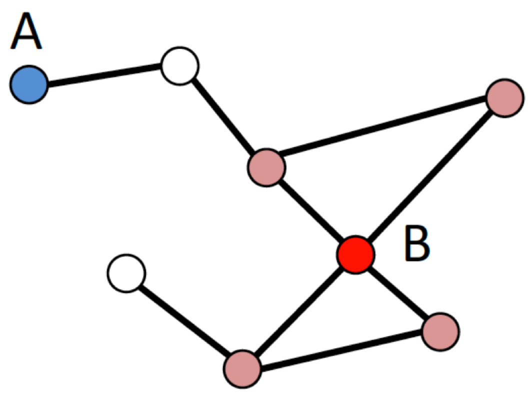
CENTRALITY

Which nodes are important based on their network?

DEGREE CENTRALITY – LOCAL MEASURE

Degree Centrality

The degree centrality of a node, v, is the degree of that node.

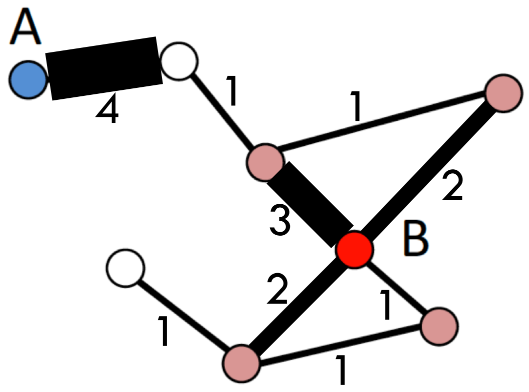


$k_A = 1$

$k_B = 4$

Weighted Degree Centrality

The weighted degree centrality of a node, v, is the sum of the weights of the incident edges.



$w_A = 4$

$w_B = 3+2+2+1$
$= 8$

MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

WHAT DOES DEGREE CENTRALITY MEAN?

1. Can a node have relatively low degree but high weighted degree? How?
2. What does degree centrality mean in the context of a movie network?
3. What does weighted degree centrality mean in the context of a movie network?
4. Does degree centrality make a character important? Why or why not?

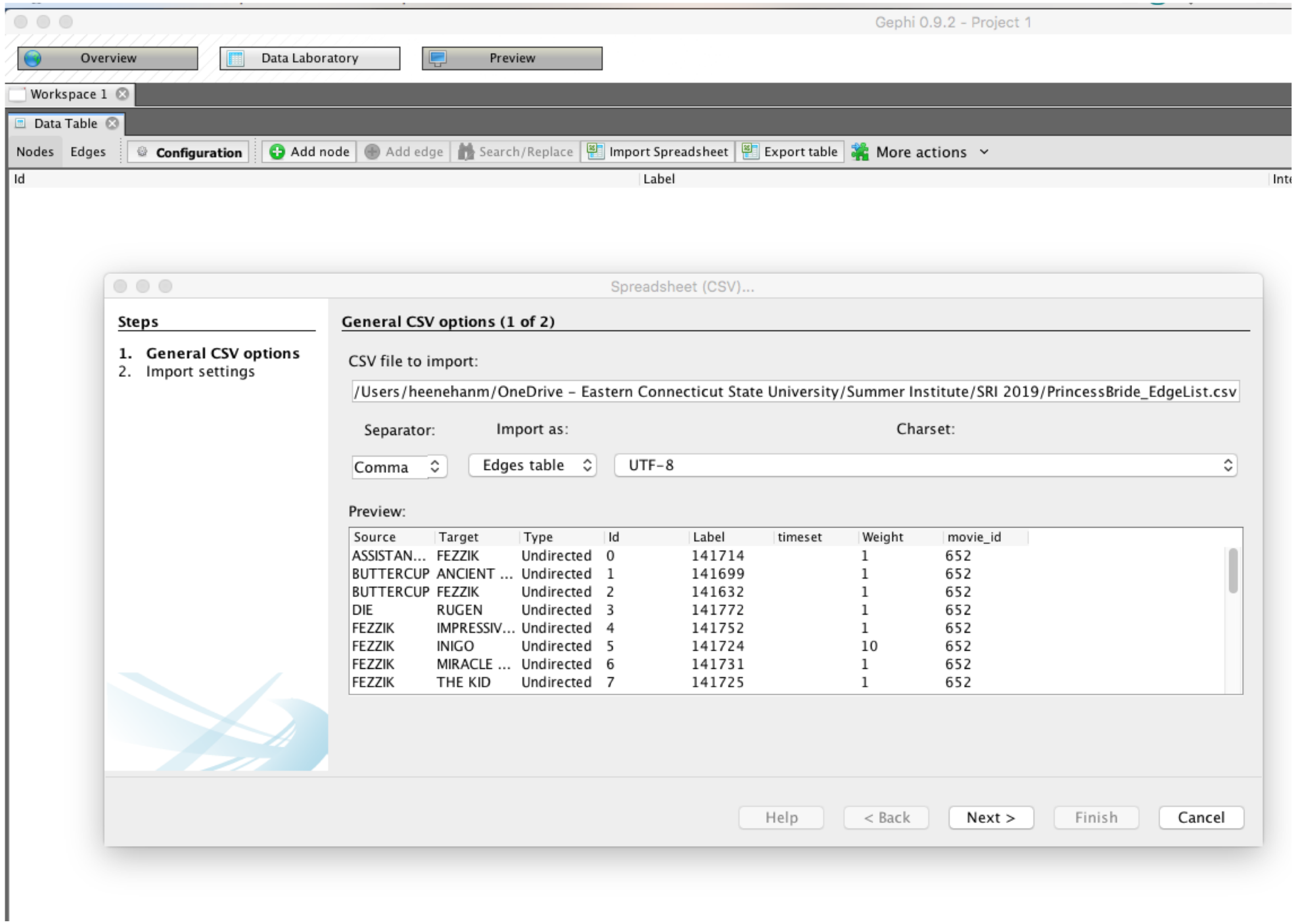
THE PRINCESS BRIDE

	Albino	Ancient Booer	Assistant Brute	Buttercup	Die	Fezzik	Grandfather	Humperdinck	Impressive Clergyman	Inigo	King	Man in Black	Miracle Max	Mother	Queen	Rugen	The Kid	Valerie	Vizzini	Westley	Yellin
Albino	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Ancient Booer	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Assistant Brute	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Buttercup	0	1	0	0	0	1	3	6	1	3	1	3	0	0	1	1	2	0	3	9	1
Die	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
Fezzik	0	0	1	1	0	0	2	1	1	10	0	3	1	0	0	0	1	0	4	2	0
Grandfather	0	0	1	3	0	2	0	2	0	2	0	0	0	1	0	0	8	0	1	2	1
Humperdinck	0	0	0	6	0	1	2	0	2	1	0	1	0	0	0	4	1	0	0	5	3
Impressive Clergyman	0	0	0	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	1	1
Inigo	0	0	1	3	1	10	2	1	0	0	0	5	2	0	0	2	1	2	5	7	1
King	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Man in Black	0	0	0	3	0	3	0	1	0	5	0	0	0	0	0	0	0	0	2	0	0
Miracle Max	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	1	0	1	0
Mother	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Queen	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Rugen	0	0	0	1	1	0	0	4	0	2	0	0	0	0	0	0	0	0	0	1	0
The Kid	0	0	0	2	0	1	8	1	0	1	0	0	0	1	0	0	0	0	1	2	0
Valerie	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0
Vizzini	0	0	0	3	0	4	1	0	0	5	0	2	0	0	0	0	1	0	0	0	0
Westley	1	0	0	9	0	2	2	5	1	7	0	0	1	0	0	1	2	0	0	0	1
Yellin	0	0	1	1	0	0	1	3	1	1	0	0	0	0	0	0	0	0	0	1	0

DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

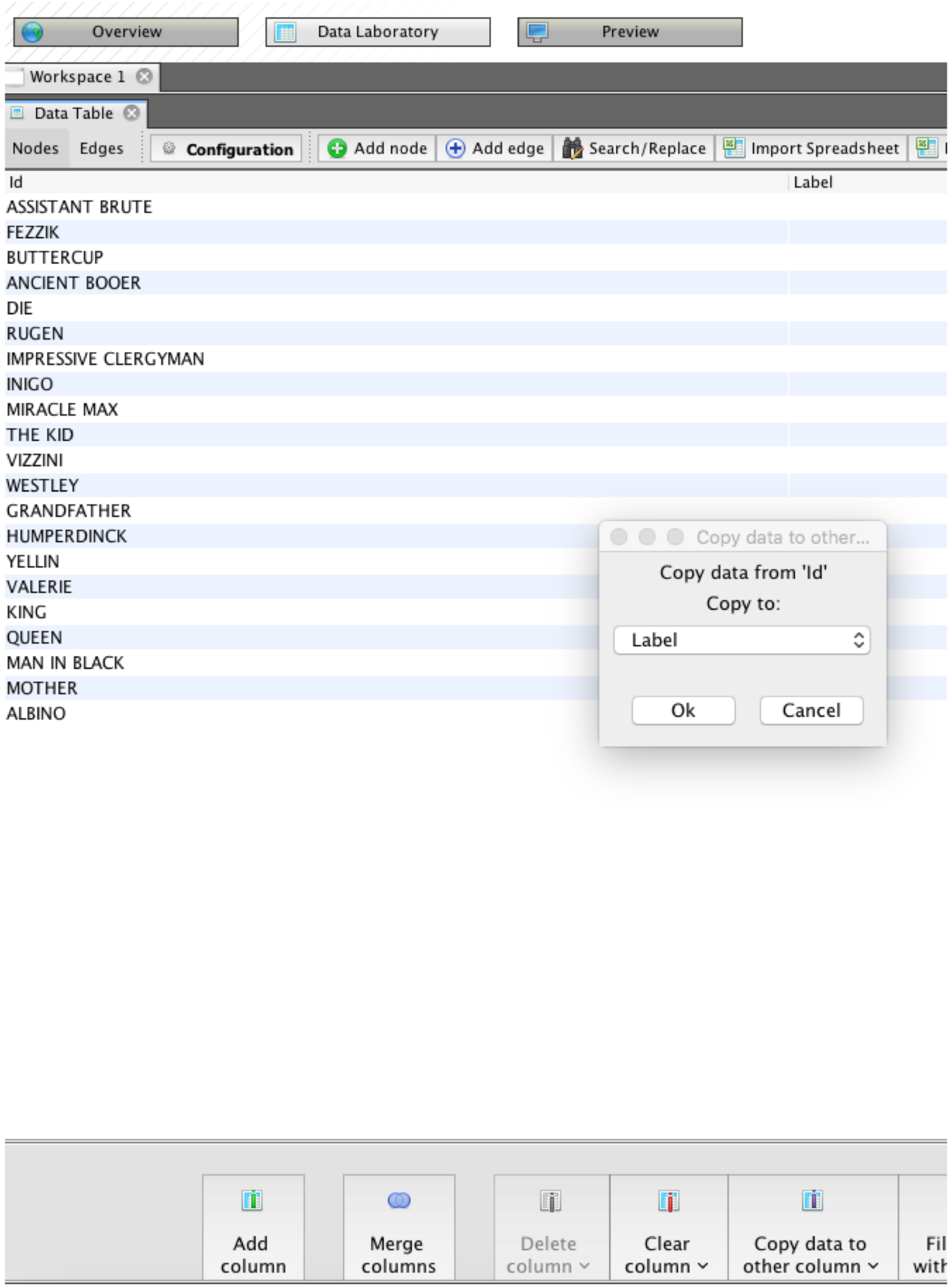
GEPHI – IMPORTING AN EDGE LIST

1. Download the file PrincessBride_EdgeList.csv
2. Open the Gephi
3. Click on "Data Laboratory"
4. Click "Import Spreadsheet"
5. Open the file PrincessBride_EdgeList.csv
   - You'll see a list of the edges. Note that, you need a "Source" and "Target" column when you import an edge list.
6. Click "Next >"
7. Click "Finish"
   - You'll get a summary with the number of nodes and edges
8. Click "OK"



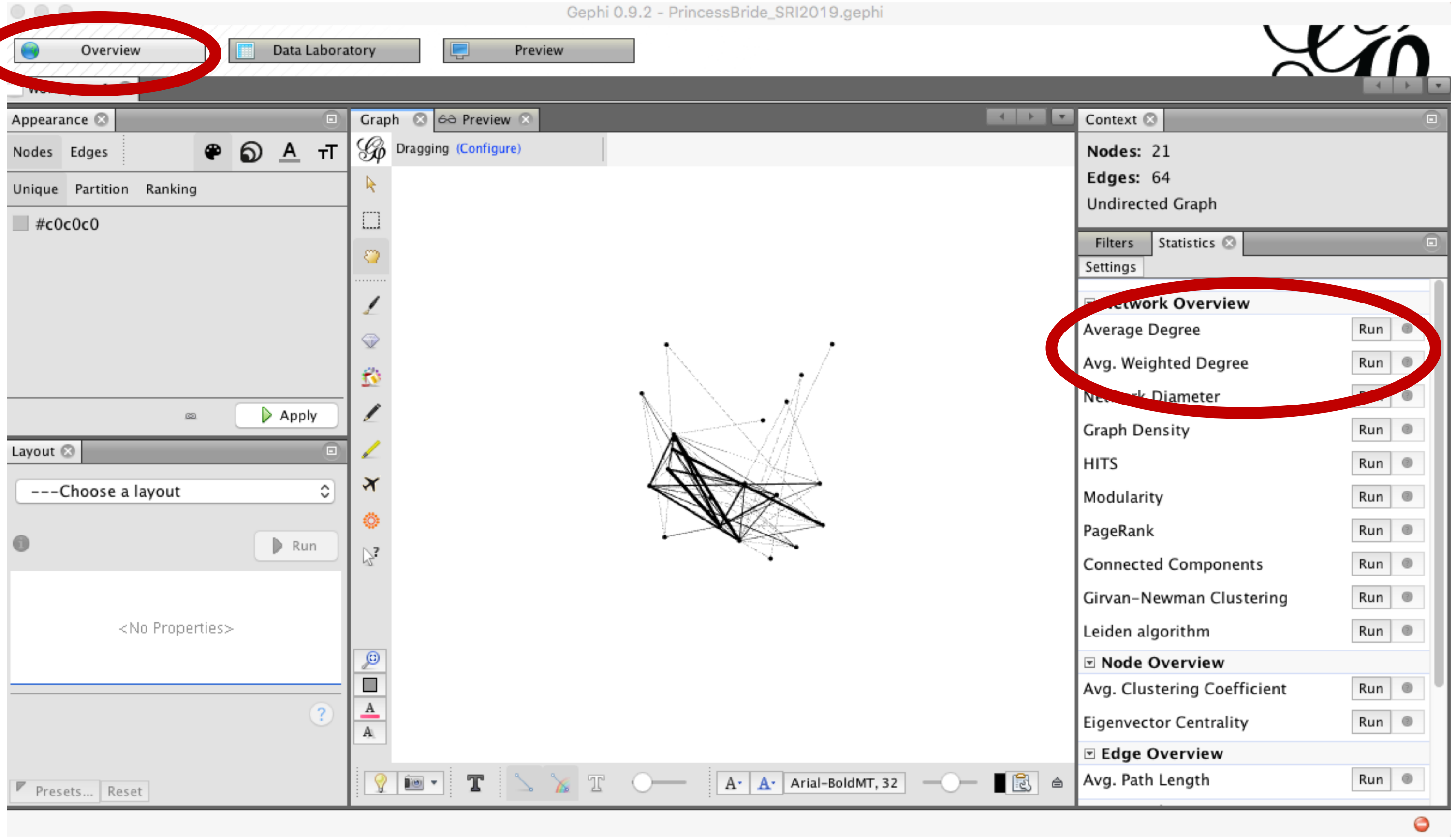
GEPHI – ADDING NODE LABELS

1. In the Data Table, click on “Nodes”
2. Notice, only the Id column is filled in. If you want to label nodes in your network, you will need to fill in the “Label” column.
3. You can copy data from one column to another.
4. Click “Copy data to other column” at the bottom of the window.
5. In the drop down menu chose “Id.”
6. In the next drop down menu choose “Label” then click “OK.”

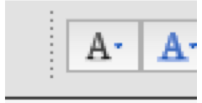


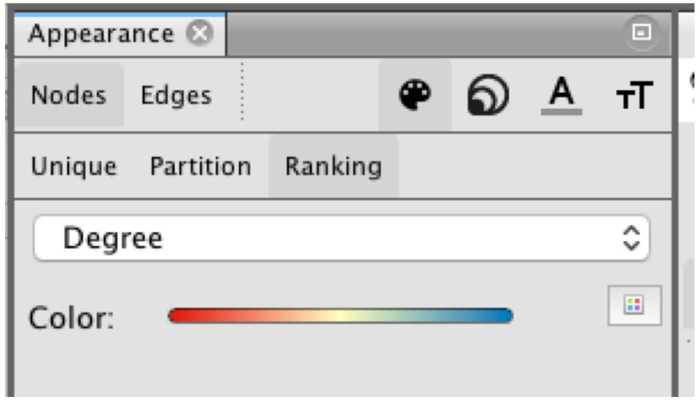
GEPHI – DEGREE CENTRALITY AND VISUALIZATION

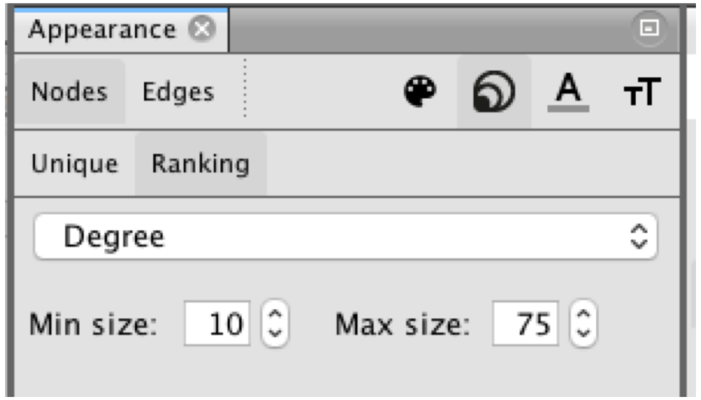
1. Click “Overview” to see your network.
2. Under the Statistics panel click “Run” next to Average Degree.
   - This will give you a chart with the degree distribution (the number of vertices of each degree that appear in the network).
   - Close this window.
3. Under the Statistics panel click “Run” next to Avg. Weighted Degree.
   - This will give the weighted degree distribution.
   - Close this window.
4. Click on the “Data Laboratory” button.
   - Here you can see the degree and weighted degree of each character.



GEPHI – DEGREE CENTRALITY AND VISUALIZATION

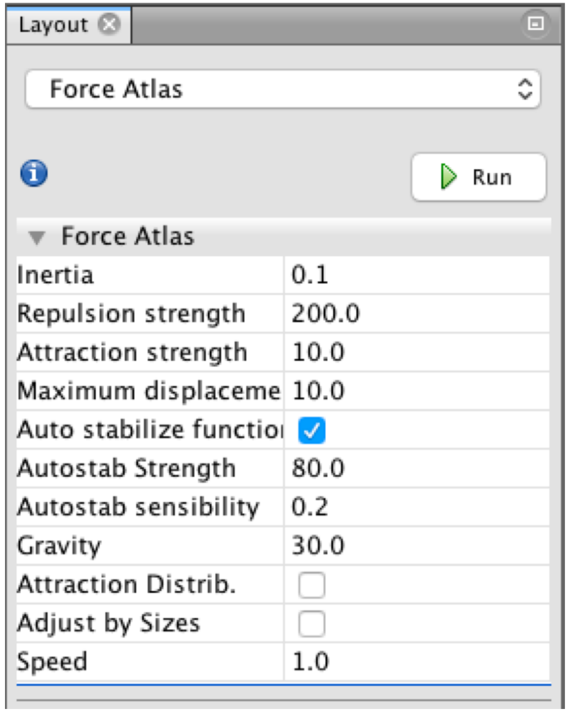
5. Go back to the "Overview." Look at the options below the graph display.
6. You can thicken the edges to more clearly see the weights using the slider: 
7. Turn Node labels on using 
   - Change their size to match the size of the node using the menu with 
8. Under "Appearance" you can rank the nodes by their degree or weighted degree, change the node size based on degree, change the colors of the nodes, etc.





GEPHI & LAYOUT

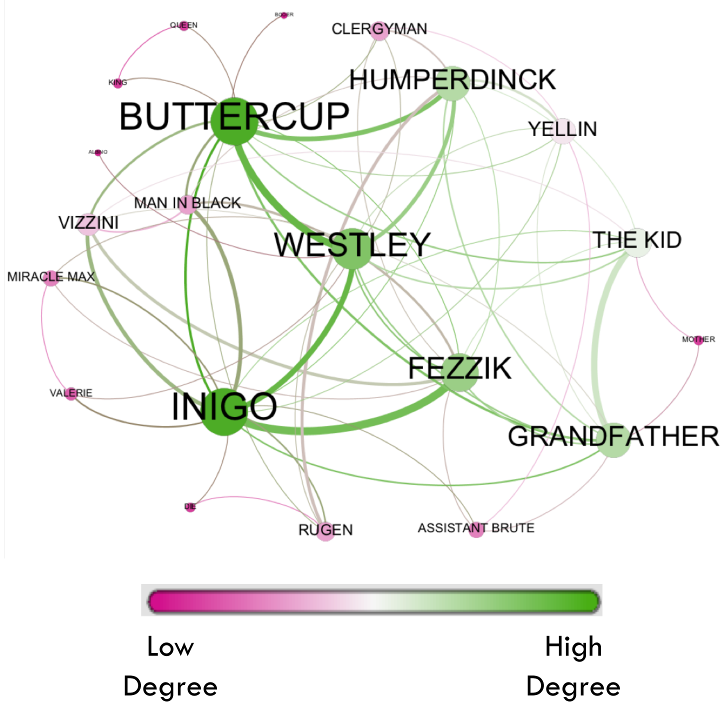
9. Go to the “Layout” tab and select “Force Atlas” from the drop down menu.
   - Idea is linked nodes attract each other and non-linked nodes repel each other.
   - Click “Run” to start the algorithm.
   - Set the “Repulsion strength” to 200,000 to expand the graph. Press enter.
   - Press ”Stop” to stop the algorithm
10. Nodes may still overlap. Check off “Adjust by Sizes” and quickly run the algorithm again.
11. To stop labels from overlapping, run the “Label Adjust” layout.



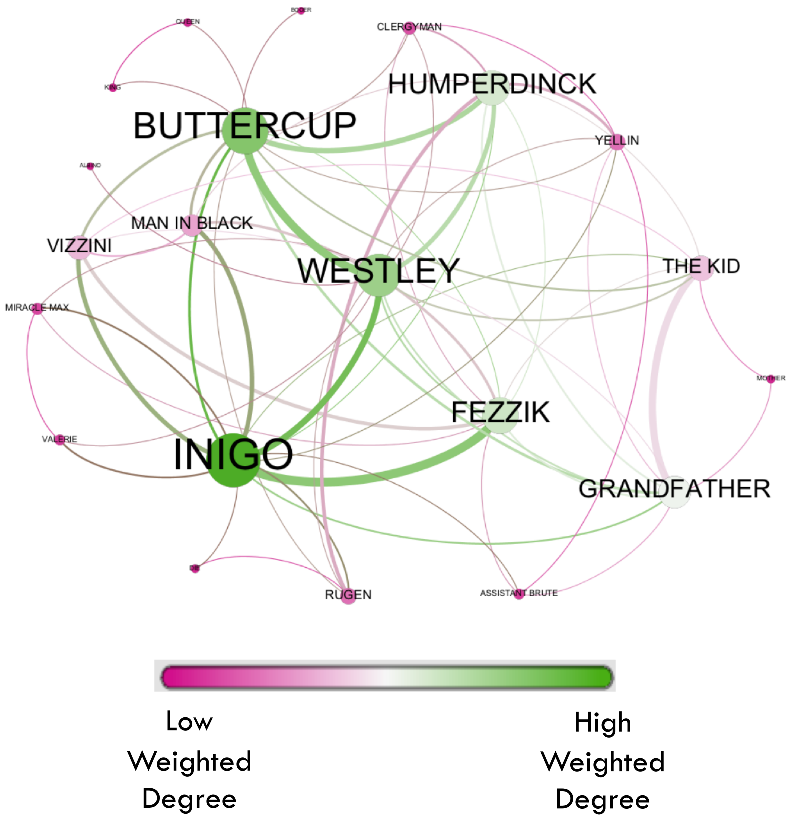
12. Experiment with different layouts. Look at the Gephi tutorial for Layouts to help you understand the parameters. Focus on ForceAtlas, ForceAtlas2, Fruchterman-Reingold, and OpenOrd https://gephi.org/users/tutorial-layouts/.
13. You can preview your graph by going to the the “Preview.” Click “Refresh” to see what the graph will look like. You can turn labels on/off, change the edges from curved to straight, etc.
14. Then you can export the graph as an svg or pdf. Alternatively, you can use the “screen shot” button at any time to get a high resolution png.

DEGREE VS. WEIGHTED DEGREE

	Degree
INIGO	14
BUTTERCUP	14
WESTLEY	12
FEZZIK	11
HUMPERDINCK	10
GRANDFATHER	10
THE KID	8
YELLIN	7
VIZZINI	6
MAN IN BLACK	5
RUGEN	5
CLERGYMAN	5
MIRACLE MAX	4
ASSISTANT BRUTE	4
VALERIE	3
DIE	2
KING	2
MOTHER	2
QUEEN	2
ALBINO	1
ANCIENT BOOER	1



	Weighted Degree
INIGO	43
BUTTERCUP	36
WESTLEY	33
FEZZIK	27
HUMPERDINCK	26
GRANDFATHER	23
THE KID	17
YELLIN	9
VIZZINI	16
MAN IN BLACK	14
RUGEN	9
ICLERGYMAN	6
MIRACLE MAX	5
ASSISTANT BRUTE	4
VALERIE	4
DIE	2
KING	2
MOTHER	2
QUEEN	2
ALBINO	1
ANCIENT BOOER	1



DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

EIGENVECTOR CENTRALITY

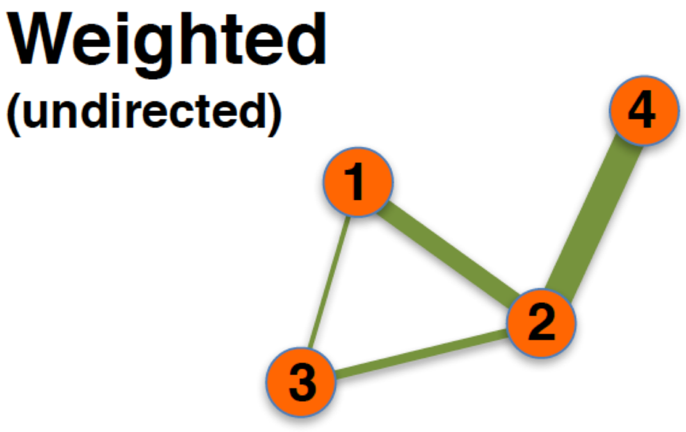
A node is important if it is connected to important nodes.

"Weighted degree centrality with a feedback loop: A [node] gets a boost for being connected to important [nodes]." (Beveridge, 2016)

The eigenvector centrality, x_i, of node i comes from solving the linear system of equations

$$x_i = \sum_{j \in V} A_{ij} x_j$$

where j is a neighbor of i.



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

A. BEVERIDGE, J. SHAN. NETWORK OF THRONES. *MATH HORIZONS*, **APRIL** (2016) 18-22.

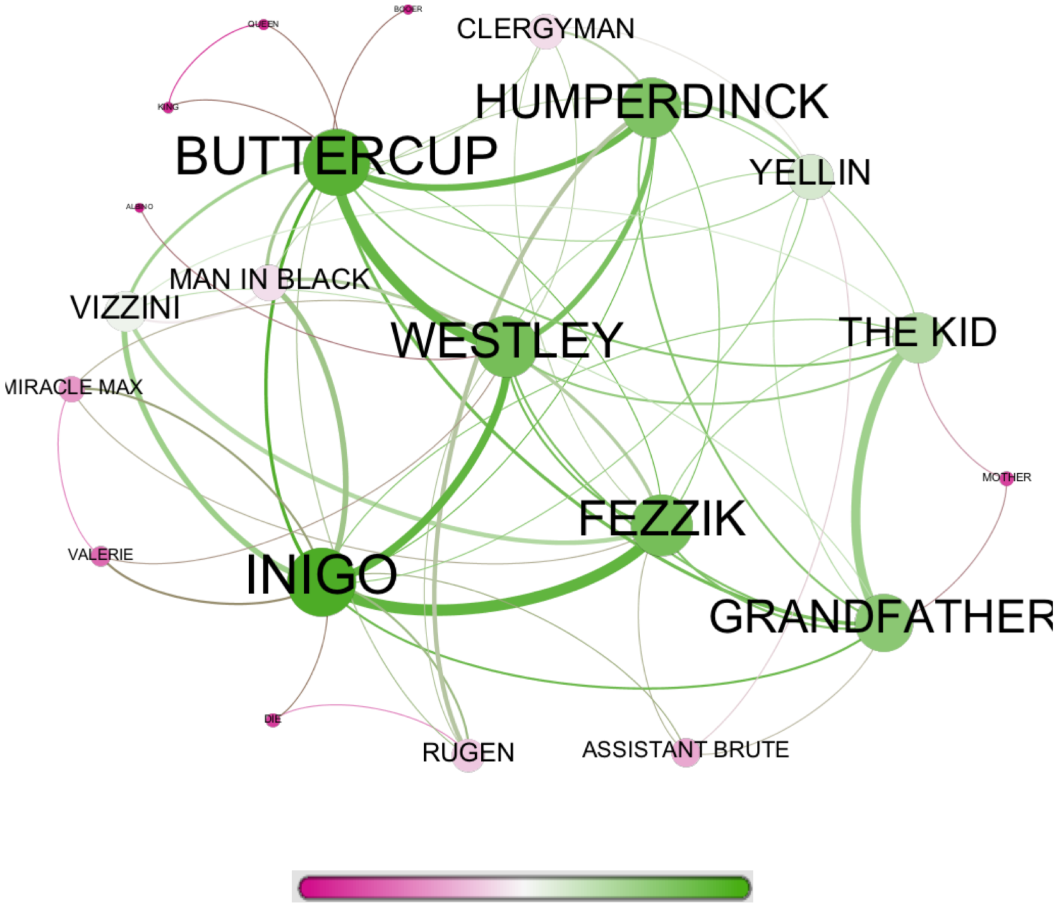
WHAT DOES EIGENVECTOR CENTRALITY MEAN?

1. What does it mean for a node in a movie network to have high eigenvector centrality?
2. Does high eigenvector centrality make a character important? Why or why not?

GEPHI & EIGENVECTOR CENTRALITY

1. Under the Statistics panel click “Run” next to Eigenvector Centrality.
   - This will give you a chart with the eigenvector centrality distribution (the number of vertices of each degree that appear in the graph).
   - Close this window.
2. Click on the “Data Laboratory” button.
   - Here you can see the eigenvector centrality of each character.
3. Under “Appearance” you can rank the nodes by their eigenvector centrality.

EIGENVECTOR CENTRALITY



Low Eigenvector Centrality

High Eigenvector Centrality

	Eigenvector Centrality
INIGO	1
BUTTERCUP	0.96795
WESTLEY	0.891104
FEZZIK	0.890407
HUMPERDINCK	0.866192
GRANDFATHER	0.834149
THE KID	0.726221
YELLIN	0.637546
VIZZINI	0.575288
MAN IN BLACK	0.502557
CLERGYMAN	0.497184
RUGEN	0.45704
ASSISTANT BRUTE	0.392665
MIRACLE MAX	0.357085
VALERIE	0.263957
MOTHER	0.18194
DIE	0.171062
KING	0.130344
QUEEN	0.130344
ANCIENT BOOER	0.113701
ALBINO	0.104436

DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

COMPARISON

	Degree Centrality	Weighted Degree Centrality	Eigenvector Centrality
INIGO	1	1	1
BUTTERCUP	1	2	2
WESTLEY	3	3	3
FEZZIK	4	4	4
HUMPERDINCK	5	5	5
GRANDFATHER	5	6	6
THE KID	7	7	7
YELLIN	8	10	8
VIZZINI	9	8	9
MAN IN BLACK	10	9	10
RUGEN	10	10	12
CLERGYMAN	10	12	11
MIRACLE MAX	13	13	14
ASSISTANT BRUTE	13	14	13
VALERIE	15	14	15
KING	16	16	18
QUEEN	16	16	18
DIE	16	16	17
MOTHER	16	16	16
ANCIENT BOOER	20	20	20
ALBINO	20	20	21

DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

PAGERANK

PageRank is the idea behind the Google search engine.

"Each vertex has an inherent importance $\beta \geq 0$, along with an importance acquired from its neighbors." (Beveridge, 2016)

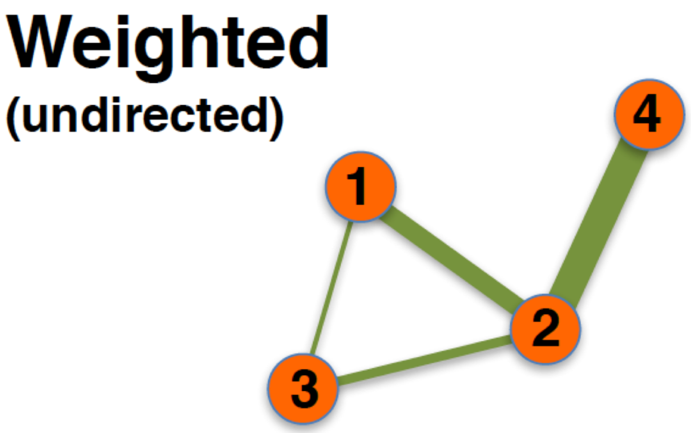
A nodes importance is divided evenly among its neighbors.

The PageRank, y_i, of vertex i is given by

$$y_i = \alpha \sum_{j \in V} \frac{A_{ij}}{k_j} y_j + \beta$$

where $\alpha + \beta = 1, \alpha, \beta \geq 0$ and j is a neighbor of i.

Researchers usually use $\beta = 0.15$.



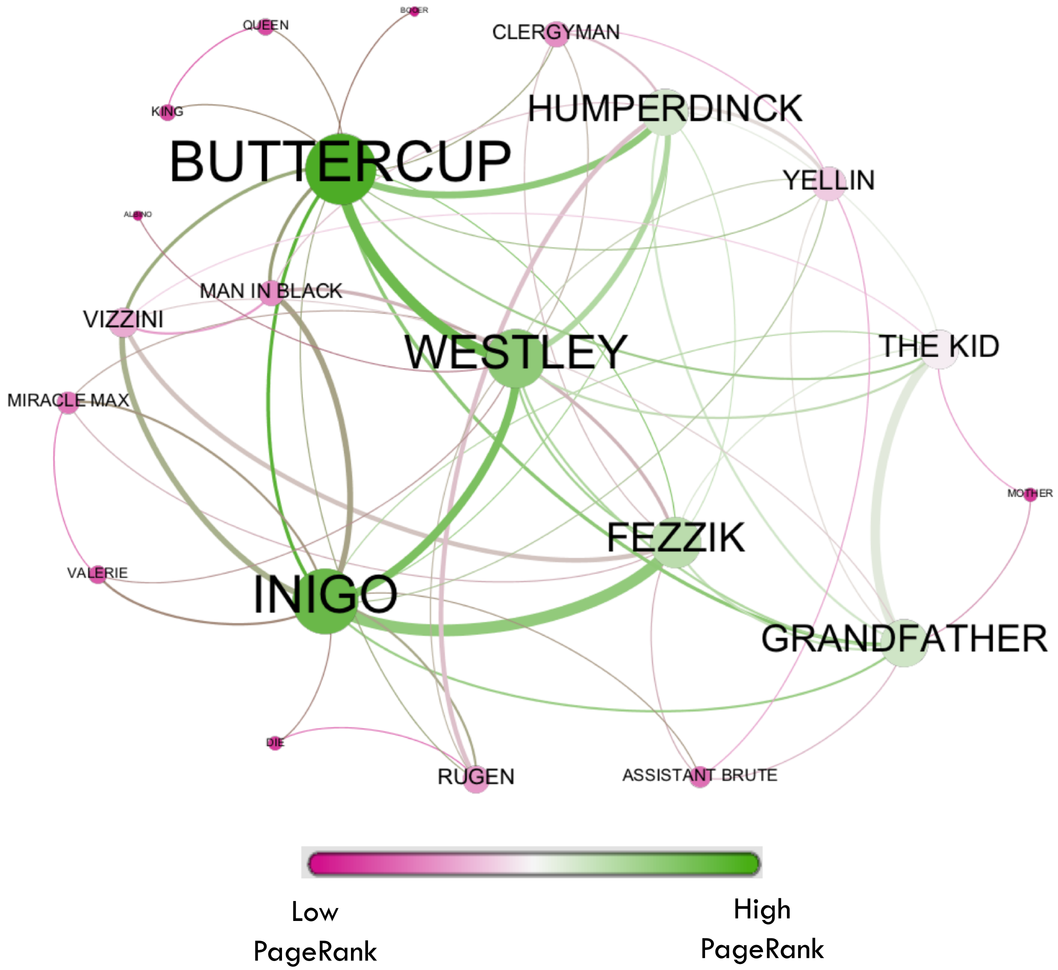
$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

A. BEVERIDGE, J. SHAN. NETWORK OF THRONES. *MATH HORIZONS*, **APRIL** (2016) 18-22.

GEPHI & PAGERANK

1. Under the Statistics panel click “Run” next to PageRank.
   - You will need to choose, p, which is like α (researches usually use α=0.85)
2. Click on the “Data Laboratory” button.
   - Here you can see the PageRank of each character.
3. Under “Appearance” you can rank the nodes by their PageRank.

PAGERANK



	PageRank
BUTTERCUP	0.108767
INIGO	0.100573
WESTLEY	0.090182
FEZZIK	0.077843
GRANDFATHER	0.072204
HUMPERDINCK	0.070926
THE KID	0.059063
YELLIN	0.051667
VIZZINI	0.044783
RUGEN	0.040841
CLERGYMAN	0.038447
MAN IN BLACK	0.038235
MIRACLE MAX	0.033206
ASSISTANT BRUTE	0.031673
VALERIE	0.026687
KING	0.023938
QUEEN	0.023938
DIE	0.020194
MOTHER	0.019558
ANCIENT BOOER	0.013748
ALBINO	0.013527

DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

COMPARISON

	Degree Centrality	Weighted Degree Centrality	Eigenvector Centrality	PageRank
INIGO	1	1	1	2
BUTTERCUP	1	2	2	1
WESTLEY	3	3	3	3
FEZZIK	4	4	4	4
HUMPERDINCK	5	5	5	6
GRANDFATHER	5	6	6	5
THE KID	7	7	7	7
YELLIN	8	10	8	8
VIZZINI	9	8	9	9
MAN IN BLACK	10	9	10	12
RUGEN	10	10	12	10
CLERGYMAN	10	12	11	11
MIRACLE MAX	13	13	14	13
ASSISTANT BRUTE	13	14	13	14
VALERIE	15	14	15	15
KING	16	16	18	16
QUEEN	16	16	18	17
DIE	16	16	17	18
MOTHER	16	16	16	19
ANCIENT BOOER	20	20	20	20
ALBINO	20	20	21	21

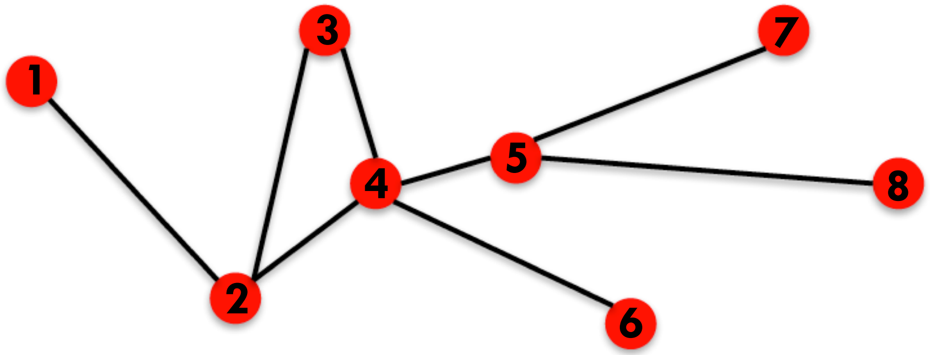
DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

PATHS AND DISTANCE

A *path* is a sequence of nodes in which each node is adjacent to the next one.

The *distance (shortest path)* between two nodes is the number of edges in the shortest path connecting them.

The distance from node i to node j is denoted d_{ij}.



Find the following distances:

- d_{15}
- d_{27}

M.E.J. NEWMAN. (2010). *NETWORKS: AN INTRODUCTION.* OXFORD UNIVERSITY PRESS.

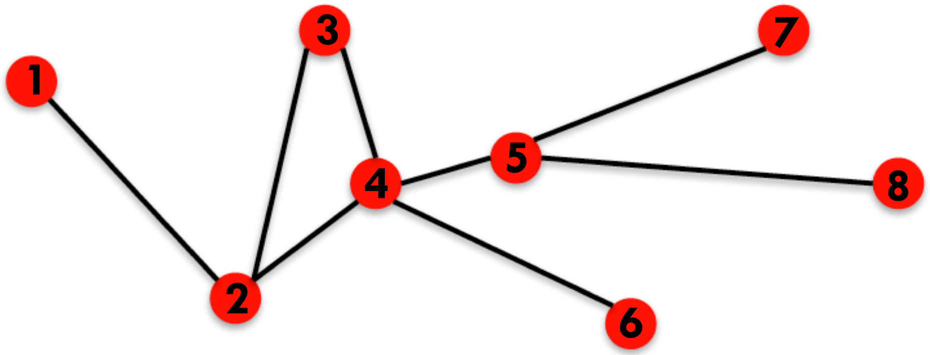
AVERAGE DISTANCE

A nodes' average distance to all other nodes is given by

$$\ell_i = \frac{1}{n} \sum_{j \in V} d_{ij}$$

where n is the number of nodes.

Find ℓ_2



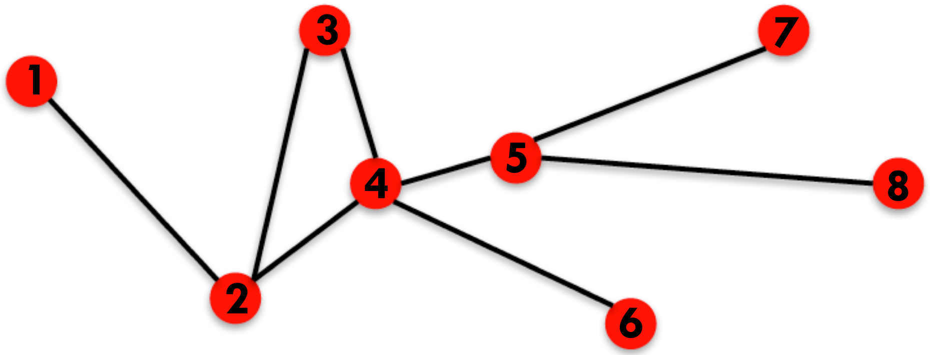
M.E.J. NEWMAN. (2010). *NETWORKS: AN INTRODUCTION.* OXFORD UNIVERSITY PRESS.

CLOSENESS CENTRALITY— GLOBAL MEASURE

The *closeness centrality* of node i is the inverse of it's average distance.

$$C_i = \frac{1}{\ell_i} = \frac{n}{\sum_{j \in V} d_{ij}}$$

Find C_2



M.E.J. NEWMAN. (2010). *NETWORKS: AN INTRODUCTION.* OXFORD UNIVERSITY PRESS.

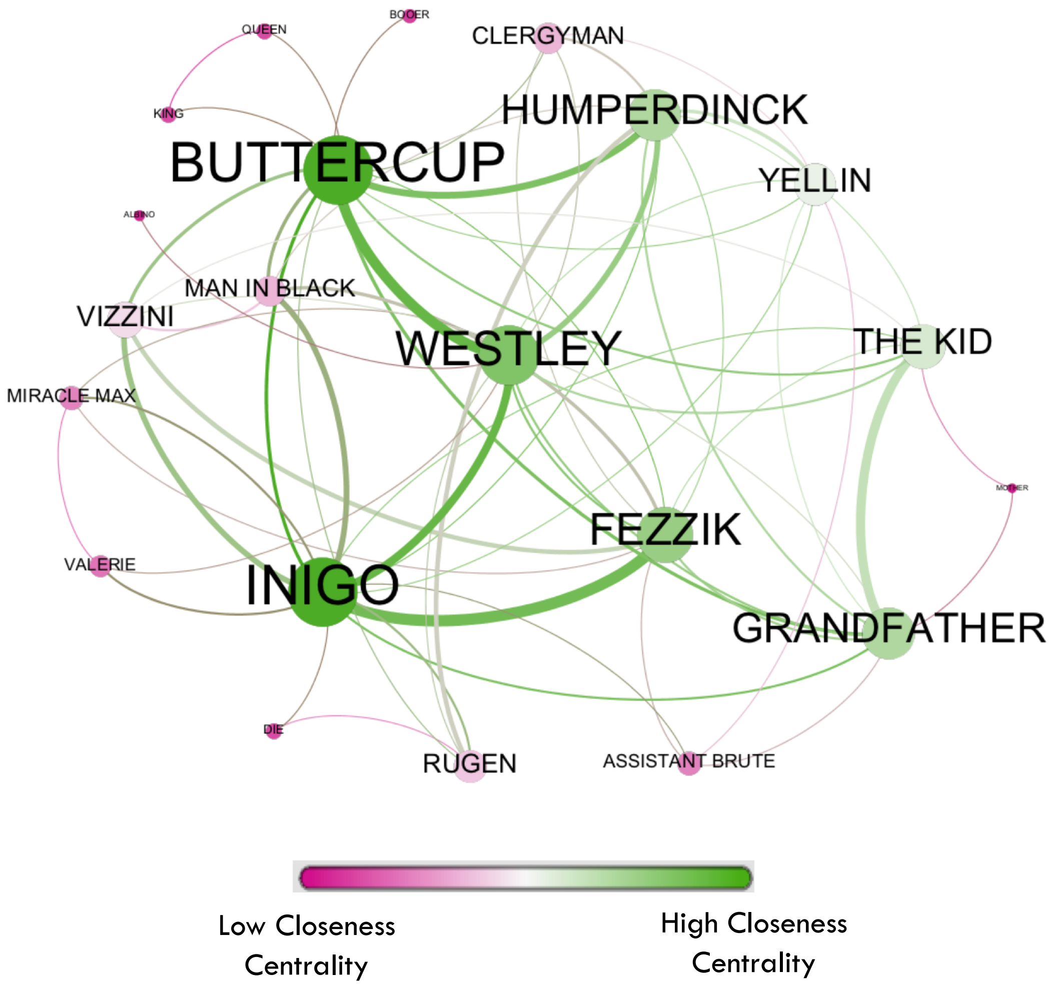
WHAT DOES CLOSENESS CENTRALITY MEAN?

1. What does it mean for a node to have high closeness centrality?
2. What does it mean for a node to have low closeness centrality?
3. What does closeness centrality mean in our movie networks?
4. Does high closeness centrality make a character important? Why or why not?

GEPHI & CLOSENESS CENTRALITY

1. Under the Statistics panel click “Run” next to Avg. Path Length. Check off “normalize” so that we can more easily compare measures.
   - This will give several centrality measures, including closeness centrality.
2. Click on the “Data Laboratory” button.
   - Here you can see the closeness centrality of each character.
3. Under “Appearance” you can rank the nodes by their closeness centrality.

CLOSENESS CENTRALITY



	Closeness Centrality
BUTTERCUP	0.769231
INIGO	0.769231
WESTLEY	0.714286
FEZZIK	0.689655
GRANDFATHER	0.666667
HUMPERDINCK	0.666667
THE KID	0.625
YELLIN	0.606061
VIZZINI	0.571429
RUGEN	0.555556
CLERGYMAN	0.540541
MAN IN BLACK	0.540541
MIRACLE MAX	0.5
ASSISTANT BRUTE	0.5
VALERIE	0.487805
KING	0.454545
QUEEN	0.454545
DIE	0.454545
ANCIENT BOOER	0.444444
ALBINO	0.425532
MOTHER	0.416667

DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

COMPARISON

	Degree Centrality	Weighted Degree Centrality	Eigenvector Centrality	PageRank	Closeness Centrality
INIGO	1	1	1	2	1
BUTTERCUP	1	2	2	1	1
WESTLEY	3	3	3	3	3
FEZZIK	4	4	4	4	4
HUMPERDINCK	5	5	5	6	6
GRANDFATHER	5	6	6	5	5
THE KID	7	7	7	7	7
YELLIN	8	10	8	8	8
VIZZINI	9	8	9	9	10
MAN IN BLACK	10	9	10	12	12
RUGEN	10	10	12	10	10
CLERGYMAN	10	12	11	11	11
MIRACLE MAX	13	13	14	13	13
ASSISTANT BRUTE	13	14	13	14	13
VALERIE	15	14	15	15	15
KING	16	16	18	16	16
QUEEN	16	16	18	17	16
DIE	16	16	17	18	16
MOTHER	16	16	16	19	21
ANCIENT BOOER	20	20	20	20	19
ALBINO	20	20	21	21	20

DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

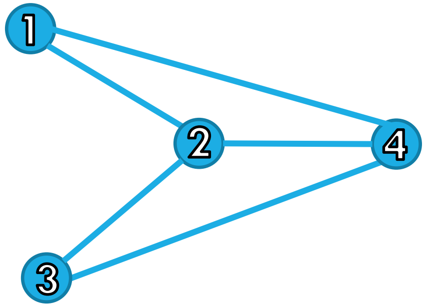
BETWEENNESS CENTRALITY – GLOBAL

The *betweenness centrality* "measures how frequently that [node] lies on short paths between other pairs of [nodes]." High betweenness means the node is a "broker of information." (Beveridge, 2016)

The betweenness, z_i, of node i is

$$z_i = \sum_{j,k \in V} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

where σ_{jk} is the number of shortest paths from j to k and $\sigma_{jk}(i)$ is the number of those paths that go through node i.



Pairs of Nodes	σ_{jk}	$\sigma_{jk}(2)$	$\frac{\sigma_{jk}(2)}{\sigma_{jk}}$
1, 2	1	1	1
1, 3	2	1	0.5
1, 4	1	0	0
2, 3	1	1	1
2, 4	1	1	1
3, 4	1	0	0
		z_2	3.5

A. BEVERIDGE, J. SHAN. NETWORK OF THRONES. *MATH HORIZONS*, **APRIL** (2016) 18-22.

WHAT DOES BETWEENNESS CENTRALITY MEAN?

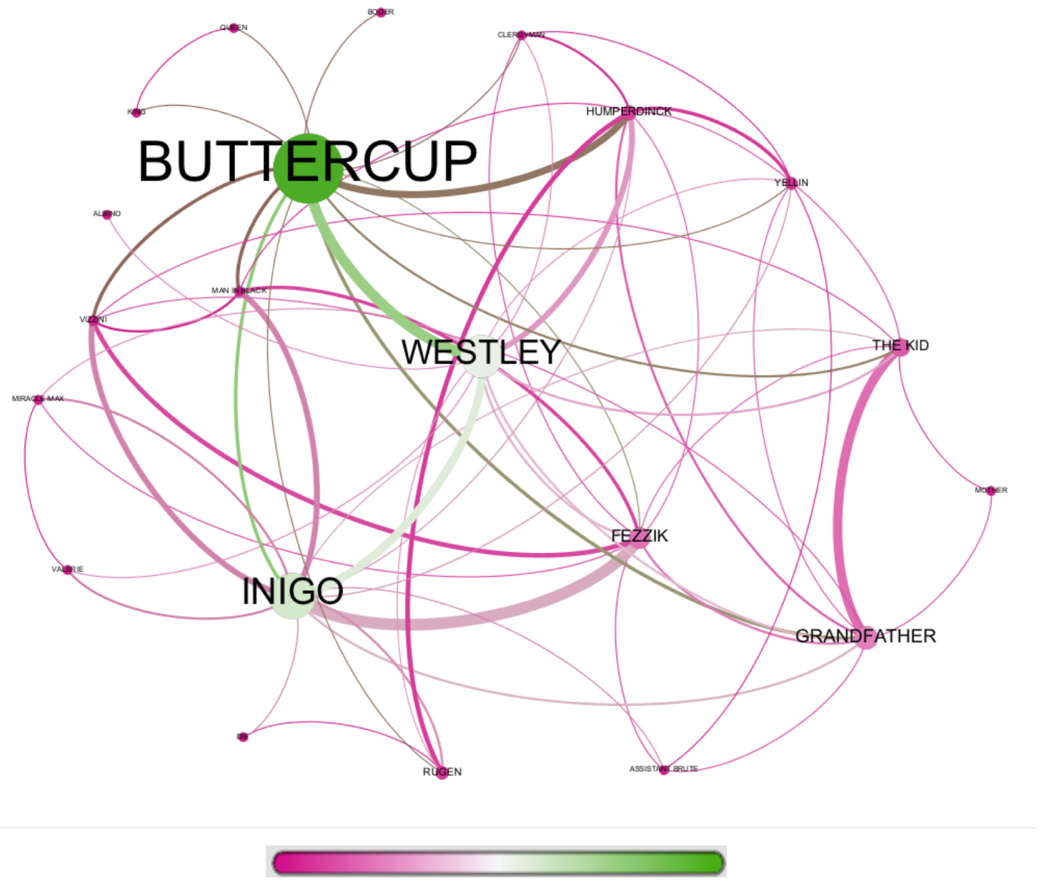
1. What does it mean for a node to have high betweenness centrality?
2. What does it mean for a node to have low betweenness centrality?
3. What does betweenness centrality mean in a movie networks?
4. Does high betweenness centrality make a character important? Why or why not?

M.E.J. NEWMAN. (2010). *NETWORKS: AN INTRODUCTION.* OXFORD UNIVERSITY PRESS.

GEPHI & BETWEENNESS CENTRALITY

1. Under the Statistics panel click “Run” next to Avg. Path Length. Check off “normalize.”
   - This will give several centrality measures, including betweenness centrality.
2. Click on the “Data Laboratory” button.
   - Here you can see the betweenness centrality of each character.
3. Under “Appearance” you can rank the nodes by their betweenness centrality.

BETWEENNESS CENTRALITY



Low Betweenness Centrality

High Betweenness Centrality

	Betweenness Centrality
BUTTERCUP	0.316044
INIGO	0.191059
WESTLEY	0.171307
GRANDFATHER	0.072185
FEZZIK	0.06095
THE KID	0.044795
HUMPERDINCK	0.027097
RUGEN	0.020395
YELLIN	0.01519
VIZZINI	0.003158
MIRACLE MAX	0.001754
MAN IN BLACK	0.000877
CLERGYMAN	0.000752
ASSISTANT BRUTE	0.000752
VALERIE	0
KING	0
QUEEN	0
DIE	0
ANCIENT BOOER	0
ALBINO	0
MOTHER	0

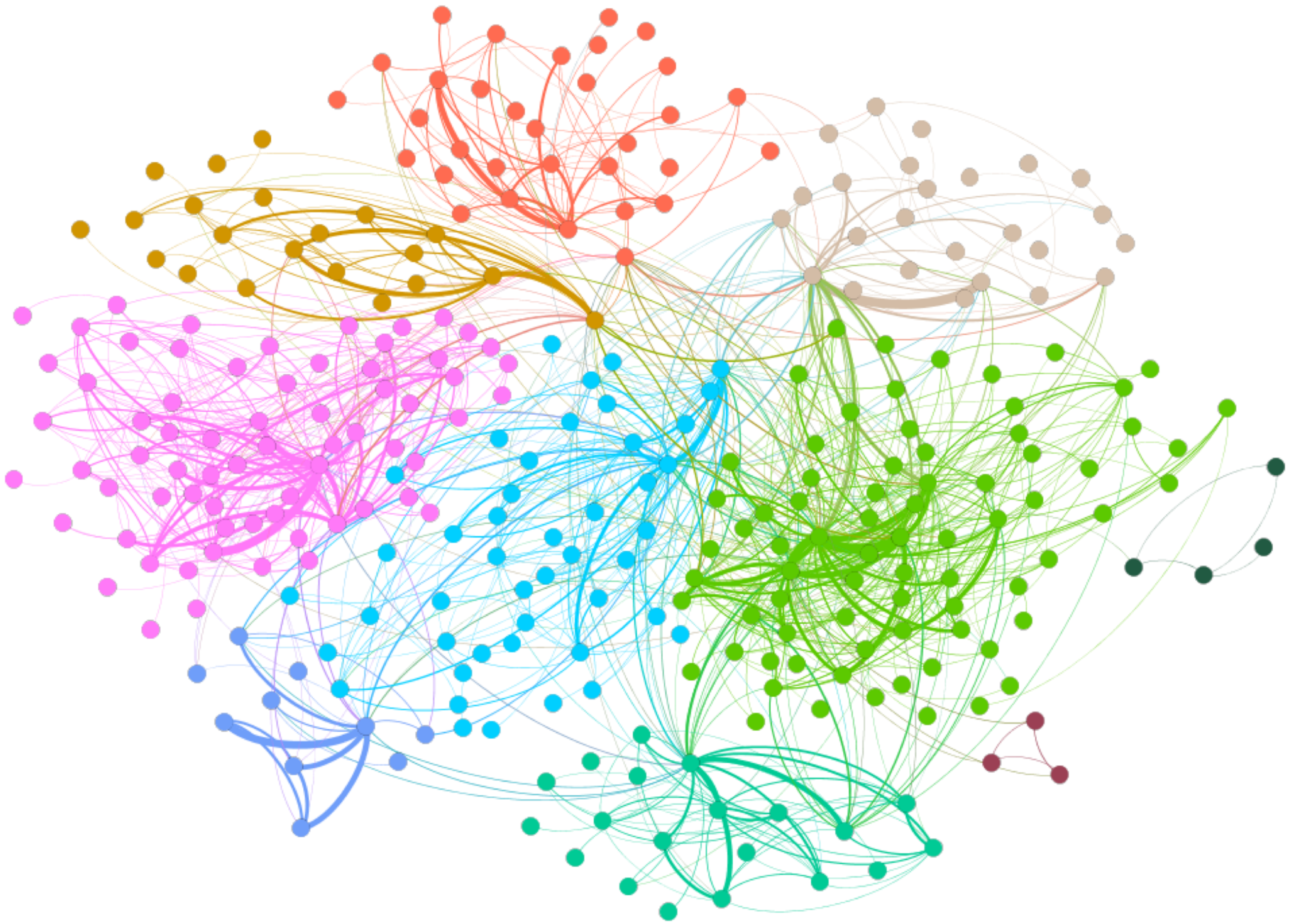
DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

COMPARISON

	Degree Centrality	Weighted Degree Centrality	Eigenvector Centrality	PageRank	Closeness Centrality	Betweenness Centrality
INIGO	1	1	1	2	1	2
BUTTERCUP	1	2	2	1	1	1
WESTLEY	3	3	3	3	3	3
FEZZIK	4	4	4	4	4	5
HUMPERDINCK	5	5	5	6	6	7
GRANDFATHER	5	6	6	5	5	4
THE KID	7	7	7	7	7	6
YELLIN	8	10	8	8	8	9
VIZZINI	9	8	9	9	10	10
MAN IN BLACK	10	9	10	12	12	12
RUGEN	10	10	12	10	10	8
CLERGYMAN	10	12	11	11	11	13
MIRACLE MAX	13	13	14	13	13	11
ASSISTANT BRUTE	13	14	13	14	13	14
VALERIE	15	14	15	15	15	15
KING	16	16	18	16	16	15
QUEEN	16	16	18	17	16	15
DIE	16	16	17	18	16	15
MOTHER	16	16	16	19	21	15
ANCIENT BOOER	20	20	20	20	19	15
ALBINO	20	20	21	21	20	15

DATA FROM KAMINSKI, J., SCHOBER, M., ALBALADEJO, R., ZASTUPAILO, O., HIDALGO, C. (2012). "MOVIEGALAXIES - SOCIAL NETWORKS IN MOVIES", HTTP://MOVIEGALAXIES.COM, AUGUST 2012, MAY 9, 2018.

WHO IS THE MOST IMPORTANT?



GAME OF THRONES – NO SPOILERS

- Fantasy series by George R. R. Martin also an HBO series
- Westeros and Essos are homes of many noble houses
- Most of the houses want to rule the kingdom
- Each house has their candidate for ruler and they are fighting for the Iron Throne
- Lots of characters
- Multiple interweaving plotlines
- Multiple locations
- Lots of drama

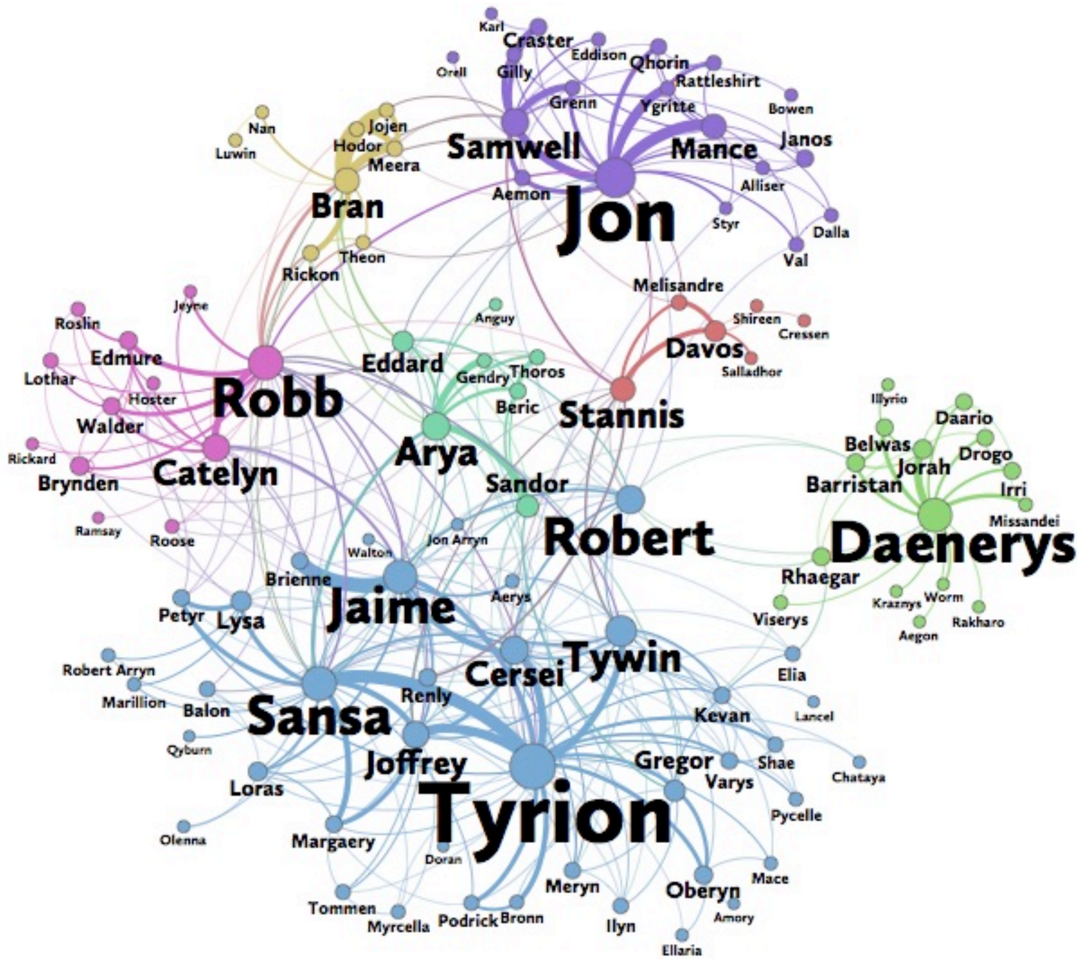
Figure 1. The *Game of Thrones* world: Westeros, the Narrow Sea, and Essos (from left to right). Sigils represent the locations of the noble houses at the beginning of the saga.

A. BEVERIDGE, J. SHAN. NETWORK OF THRONES. *MATH HORIZONS*, **APRIL** (2016) 18-22.

NETWORK OF THRONES

BY BEVERIDGE AND SHAN

- Looked at the third book “A Storm of Swords” by George R. R. Martin
- Extracted characters from the book.
- Linked two characters each time they were mentioned within 15 words of one another.
- Used network analysis “to make sense of the intricate character relationships and their bearing on the future plot.”



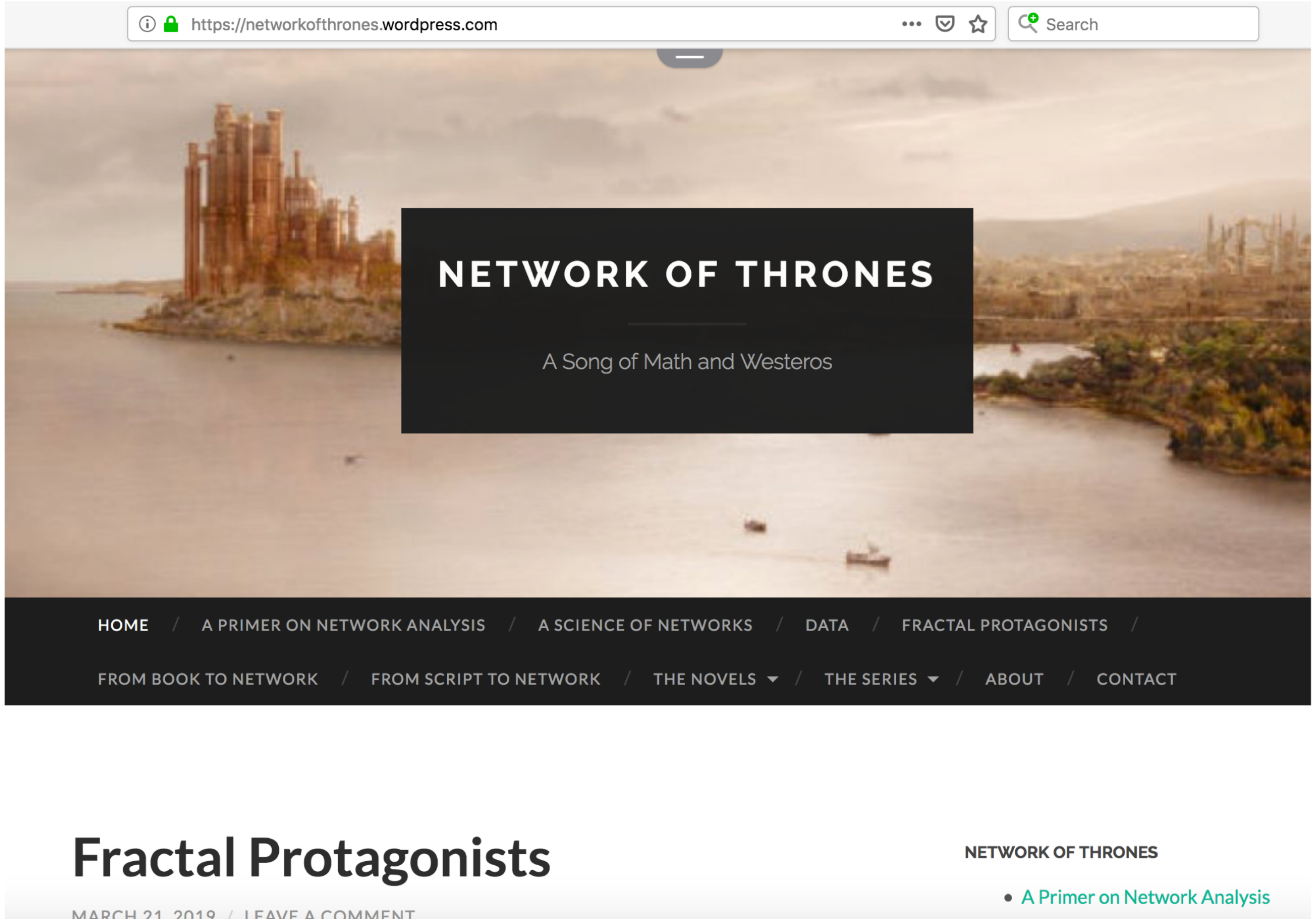
A. BEVERIDGE, J. SHAN. NETWORK OF THRONES. *MATH HORIZONS*, **APRIL** (2016) 18-22.

NETWORK & DATA

Andrew Beveridge has a blog “Network of Thrones, A Song of Math and Westeros” (https://networkofthrones.wordpress.com/)

All of the data are shared on Github:

https://github.com/mathbeveridge/asoiaf

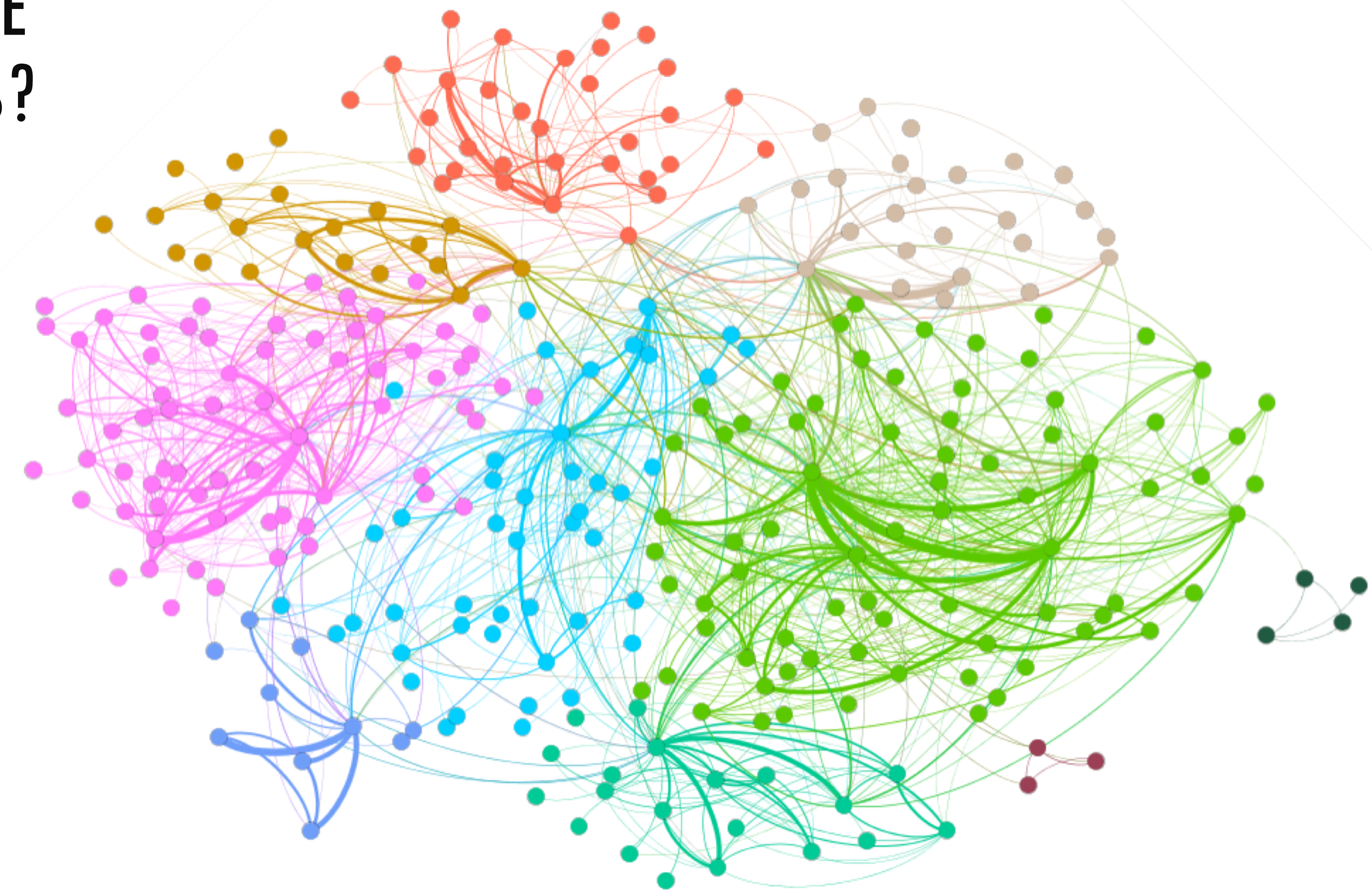


WHO WILL WIN THE GAME OF THRONES?

303 Nodes

1008 Edges

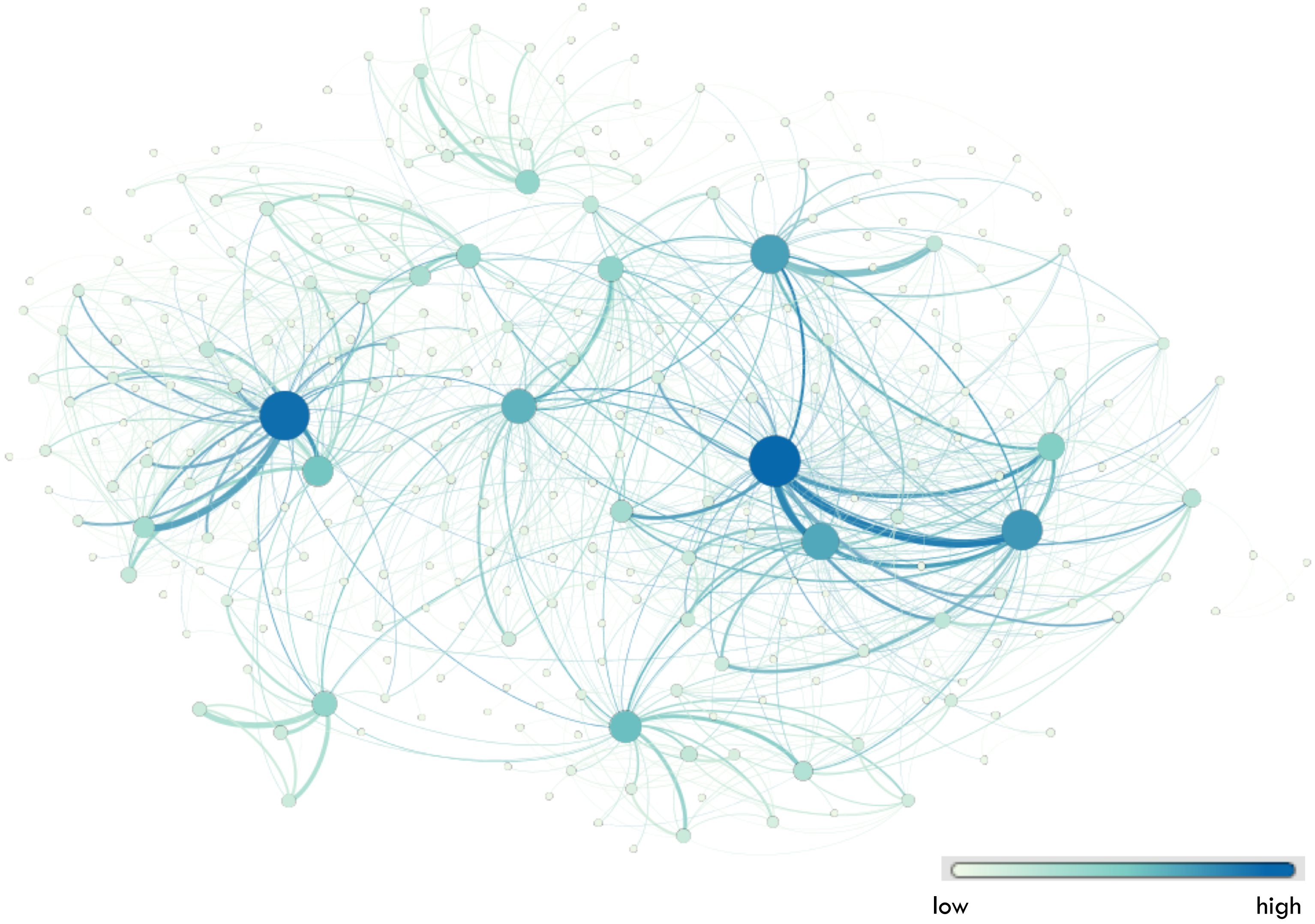
Thicker edges indicate more interactions.



DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

WHO WILL WIN THE GAME OF THRONES?

Nodes sized and colored by weighted degree.

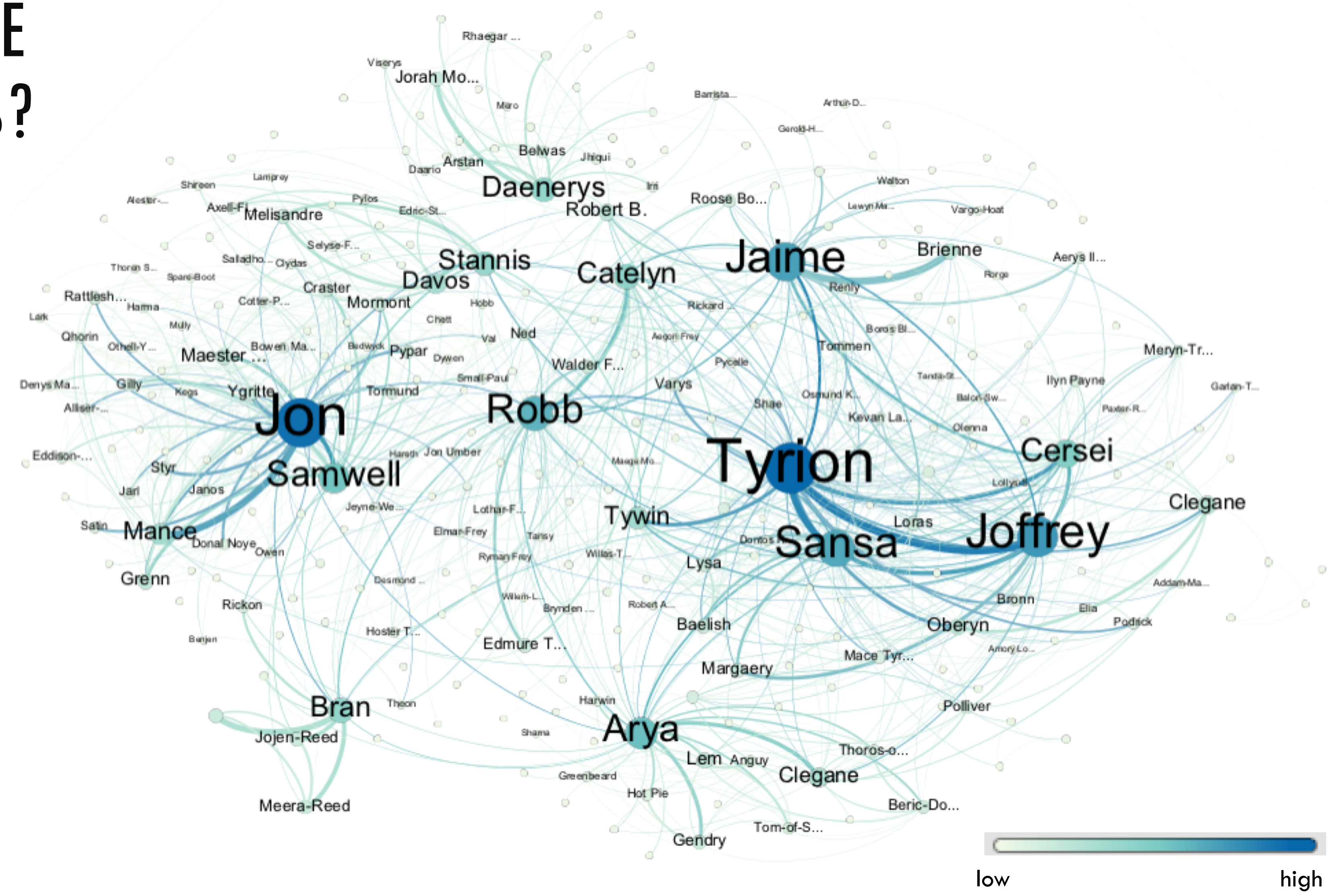


DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

WHO WILL WIN THE GAME OF THRONES?

Nodes sized and colored by weighted degree.

1. Tyrion
2. Jon Snow
3. Joffrey
4. Jaime
5. Sansa
6. Robb
7. Arya
8. Samwell
9. Cersei
10. Catelyn

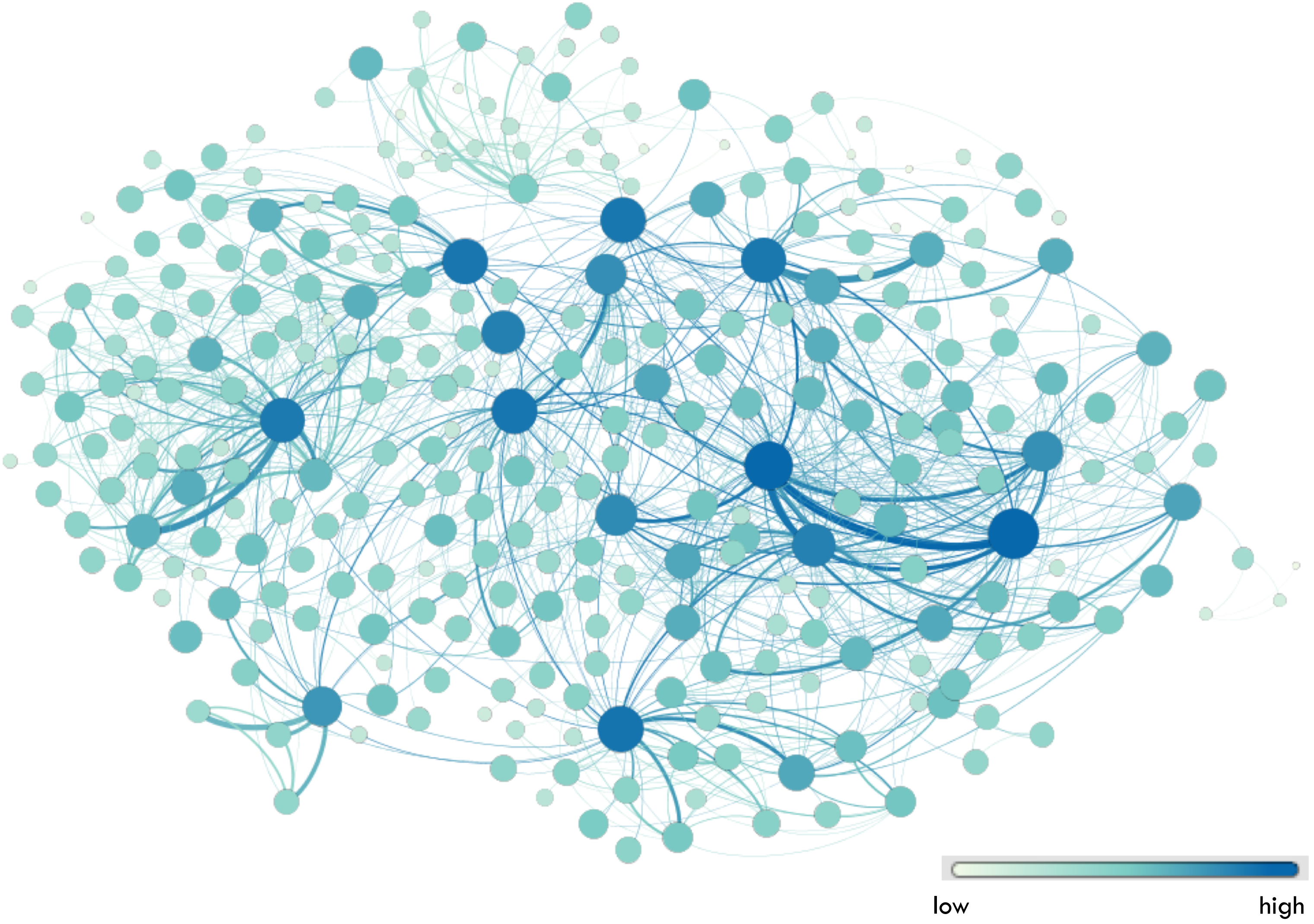


DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

WHO WILL WIN THE GAME OF THRONES?

Nodes sized and colored by closeness centrality.

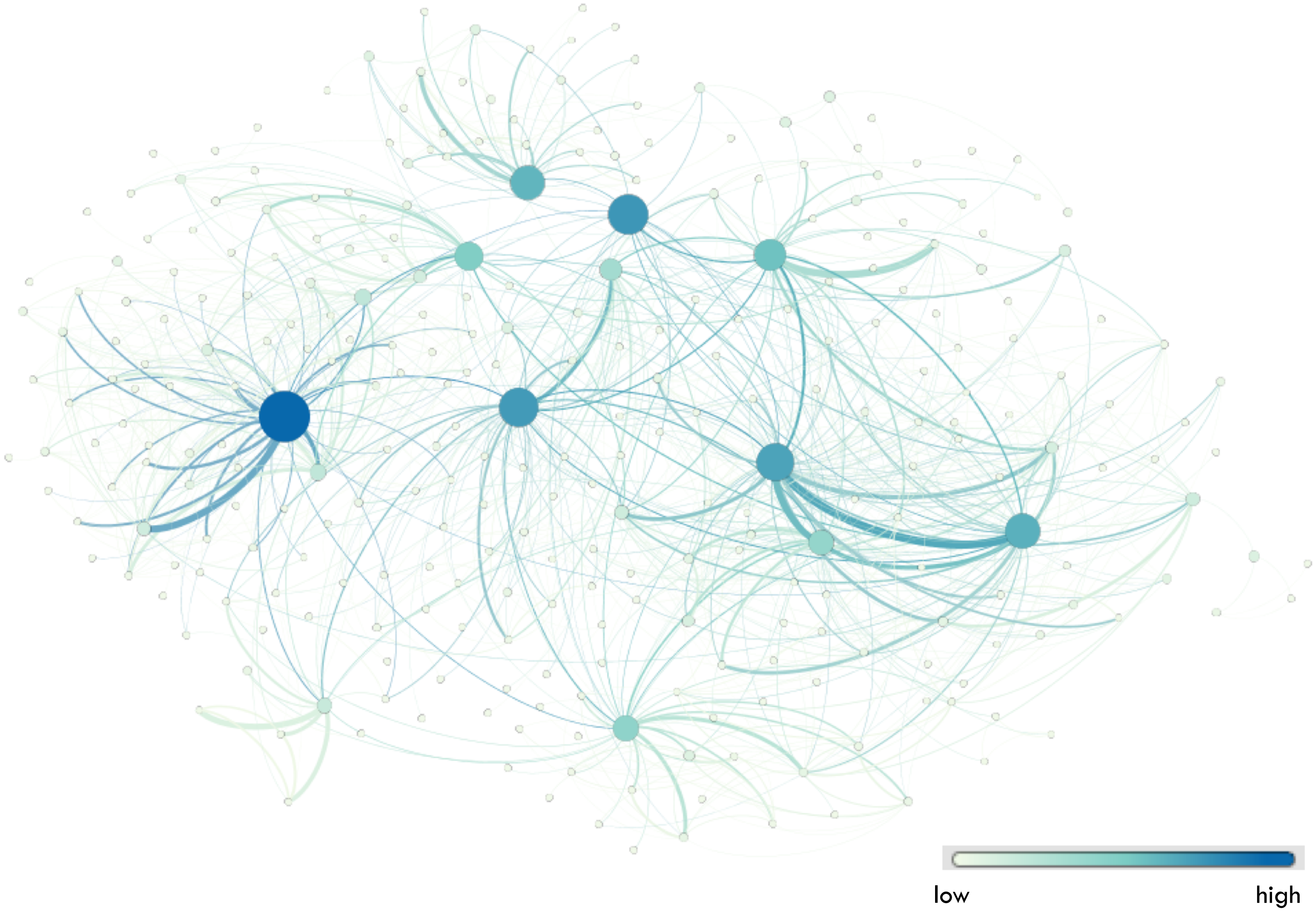
1. Joffrey
2. Tyrion
3. Arya
4. Robb
5. Stannis
6. Robert Baratheon
7. Jaime
8. Jon Snow
9. Ned Stark
10. Sansa



DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

WHO WILL WIN THE GAME OF THRONES?

Nodes sized and colored by betweenness centrality.

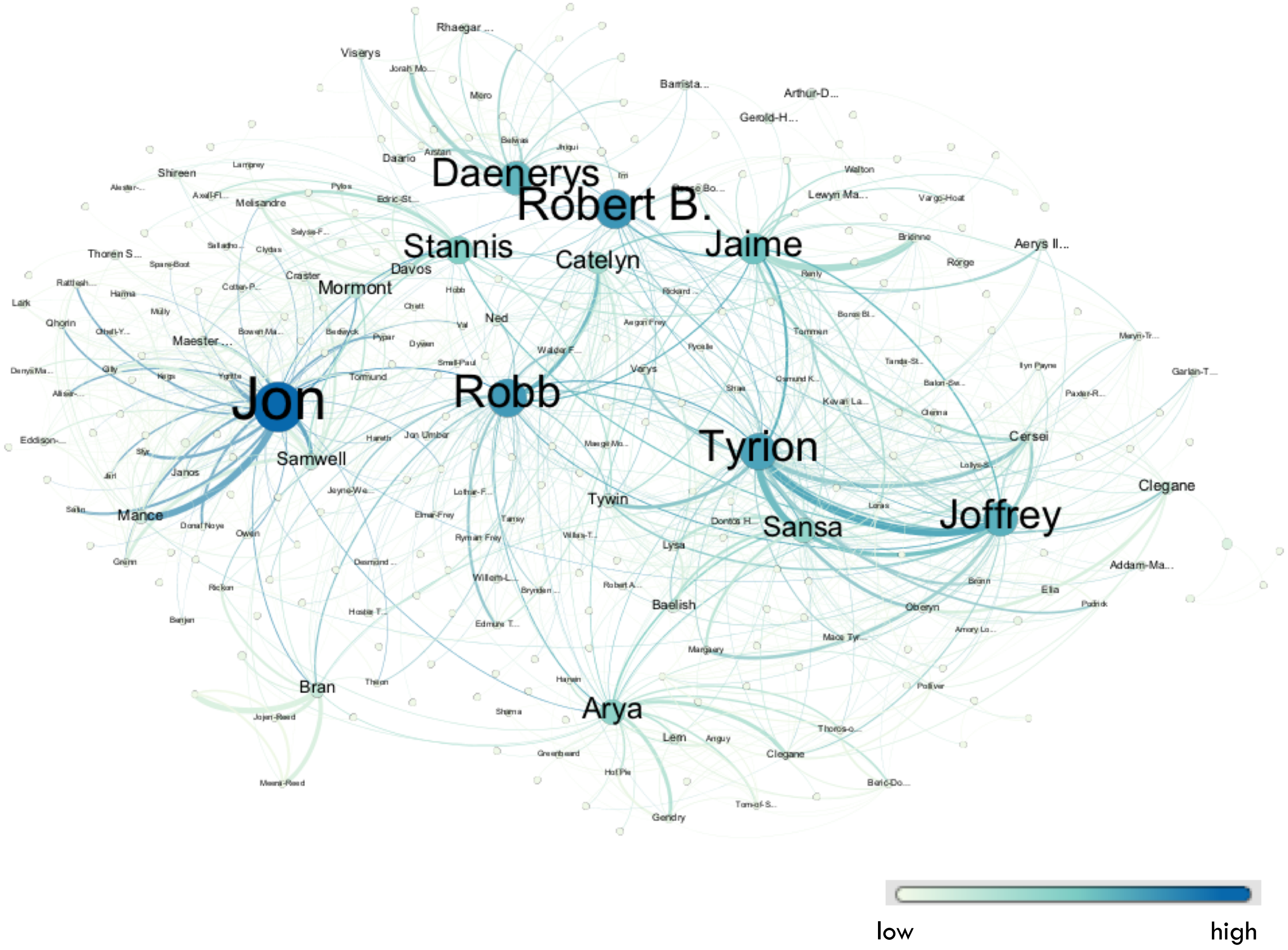


DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

WHO WILL WIN THE GAME OF THRONES?

Nodes sized and colored by betweenness centrality.

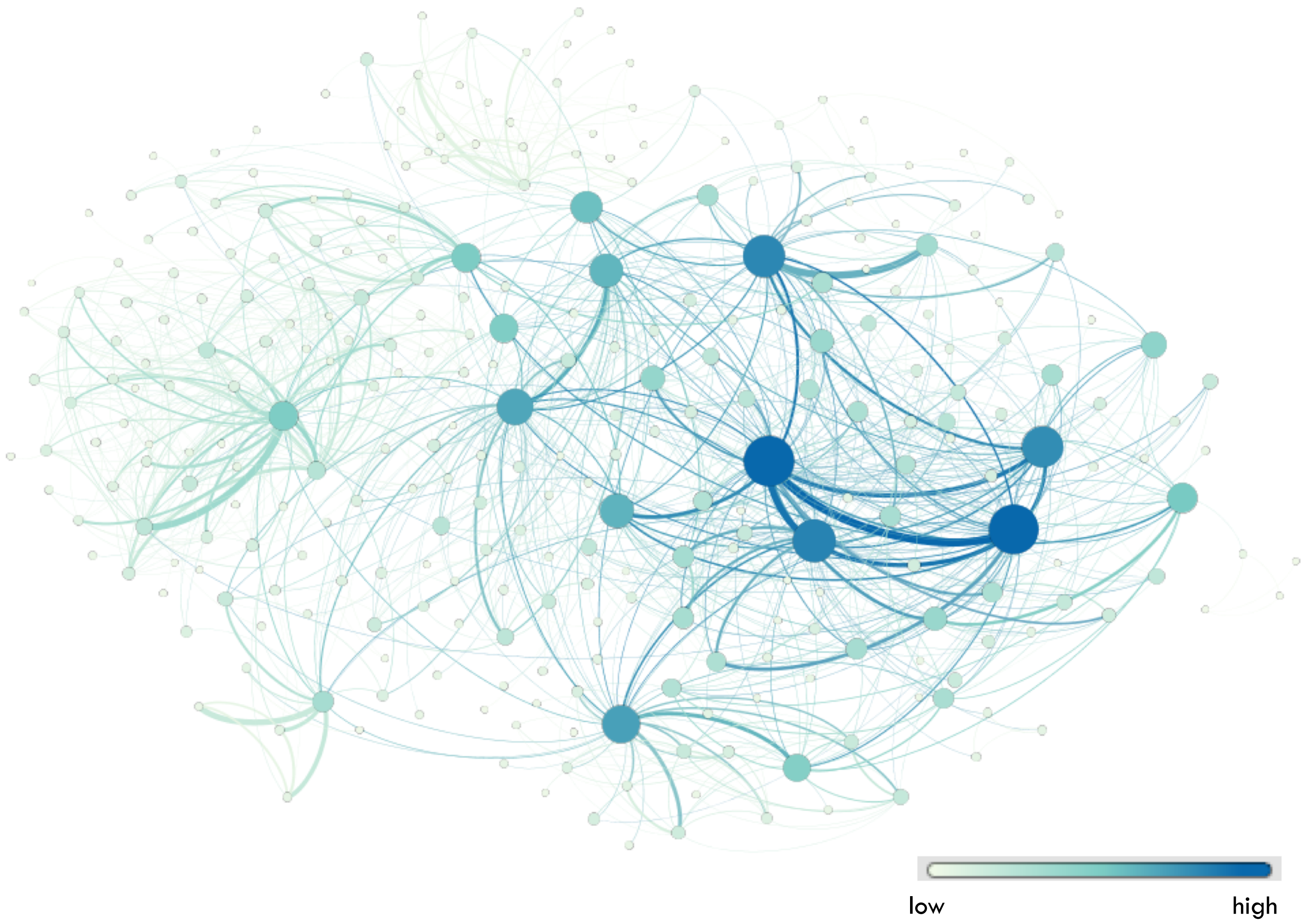
1. Jon Snow
2. Robert Baratheon
3. Robb
4. Tyrion
5. Joffrey
6. Daenerys
7. Jaime
8. Stannis
9. Arya
10. Sansa



DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

WHO WILL WIN THE GAME OF THRONES?

Nodes sized and colored by by eigenvector centrality.

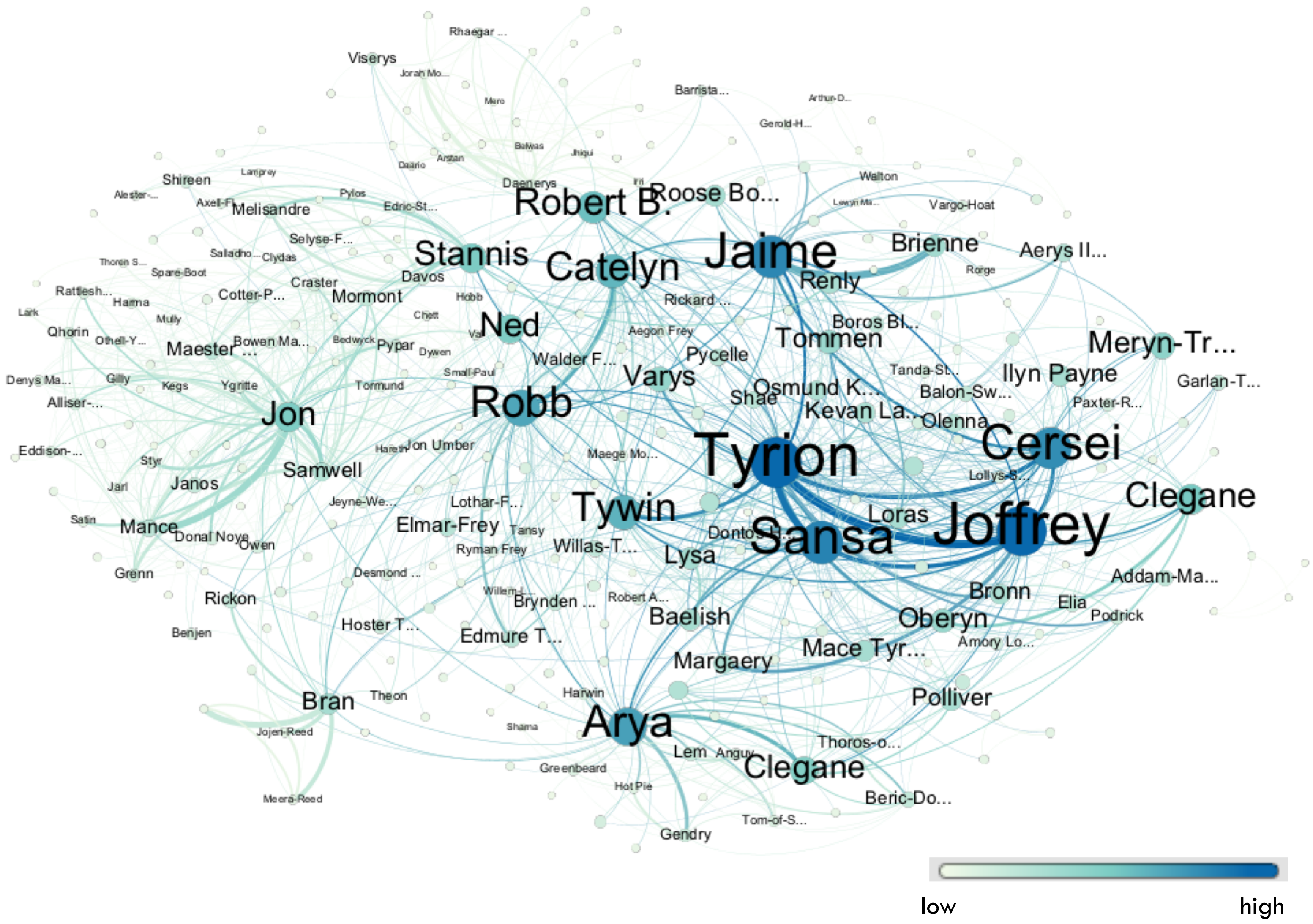


DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

WHO WILL WIN THE GAME OF THRONES?

Nodes sized and colored by eigenvector centrality.

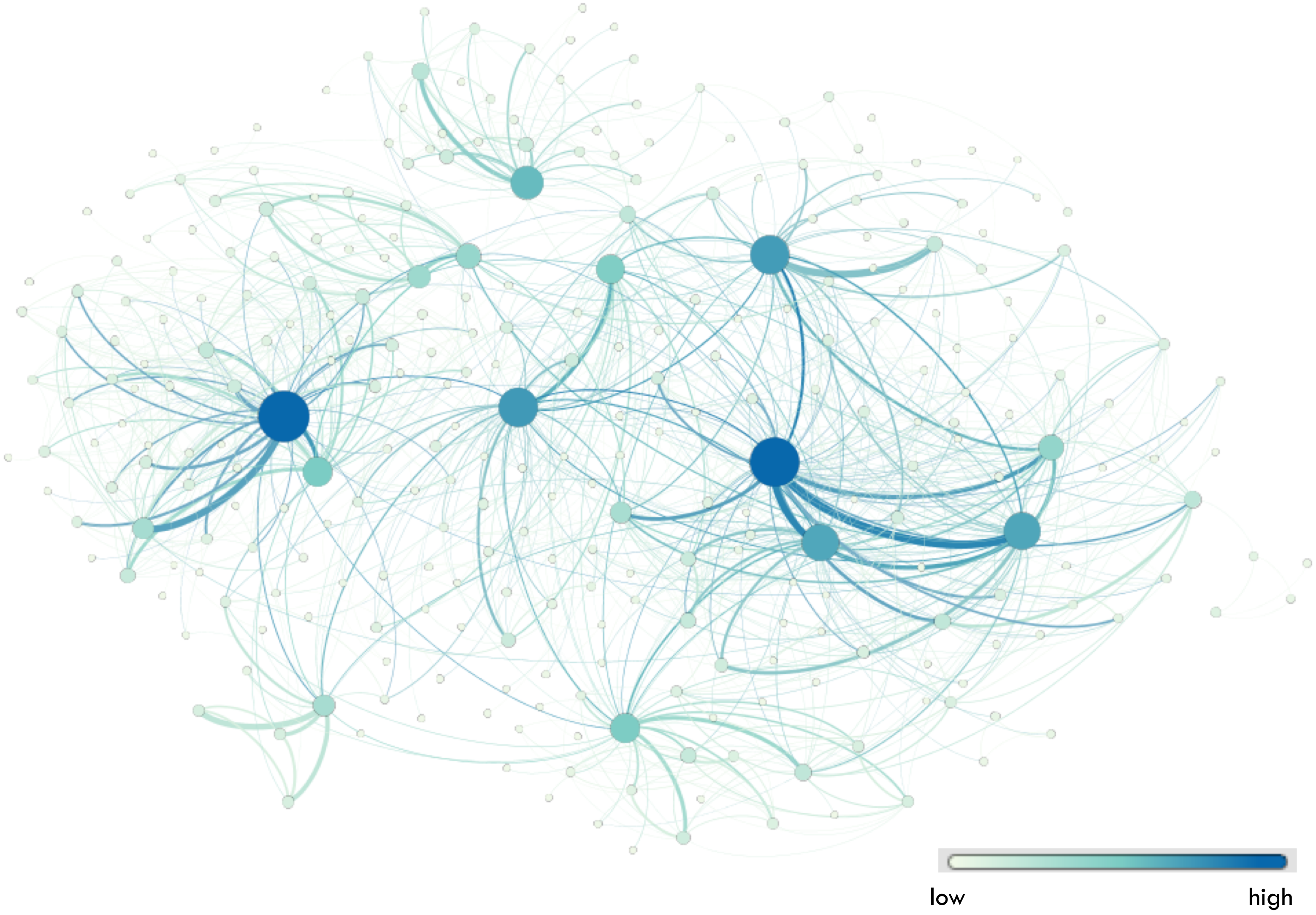
1. Tyrion
2. Joffrey
3. Sansa
4. Jaime
5. Cersei
6. Arya
7. Robb
8. Tywin Lannister
9. Catelyn
10. Robert Baratheon



DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

WHO WILL WIN THE GAME OF THRONES?

Nodes sized and colored by PageRank.

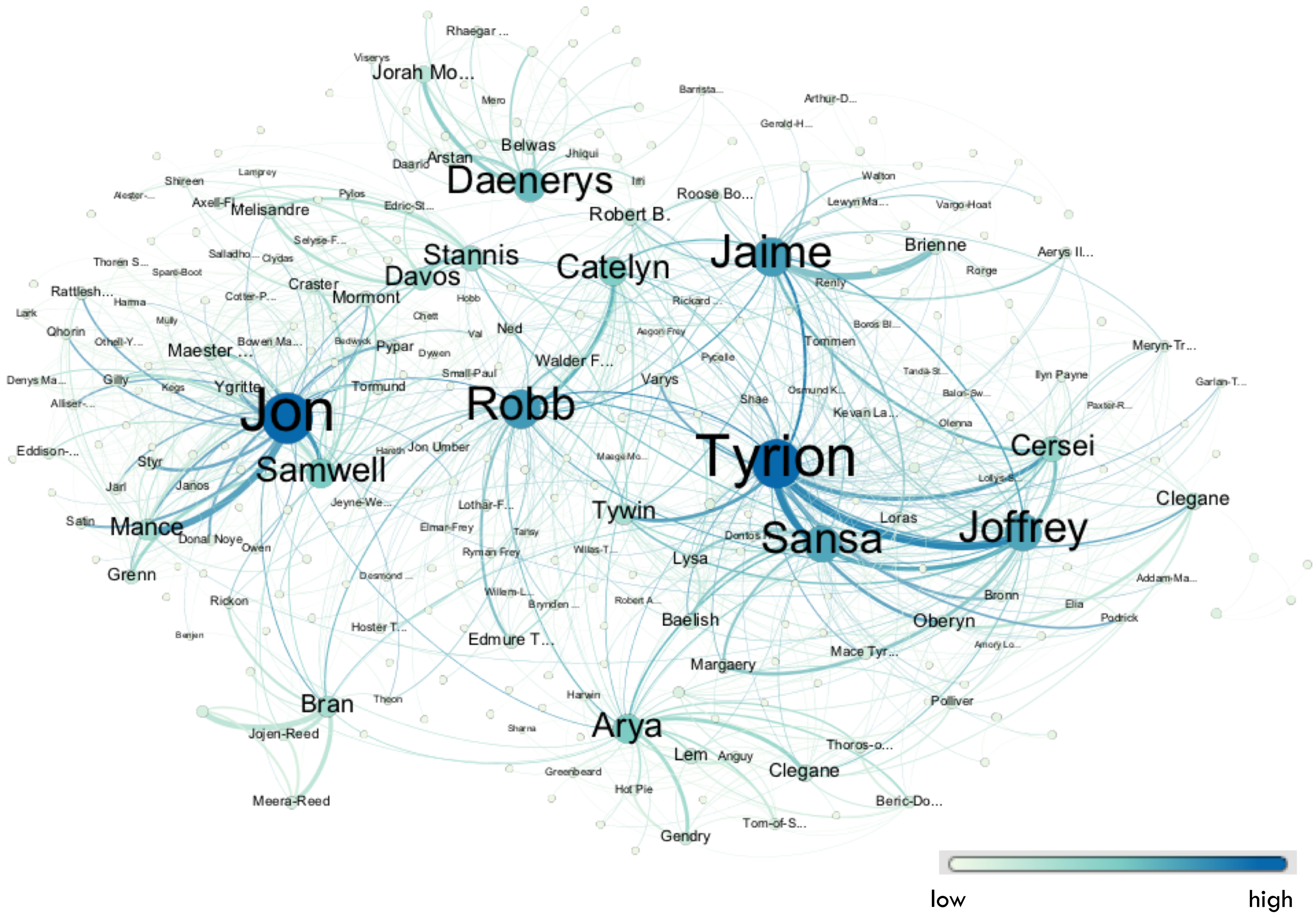


DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

WHO WILL WIN THE GAME OF THRONES?

Nodes sized and colored by PageRank.

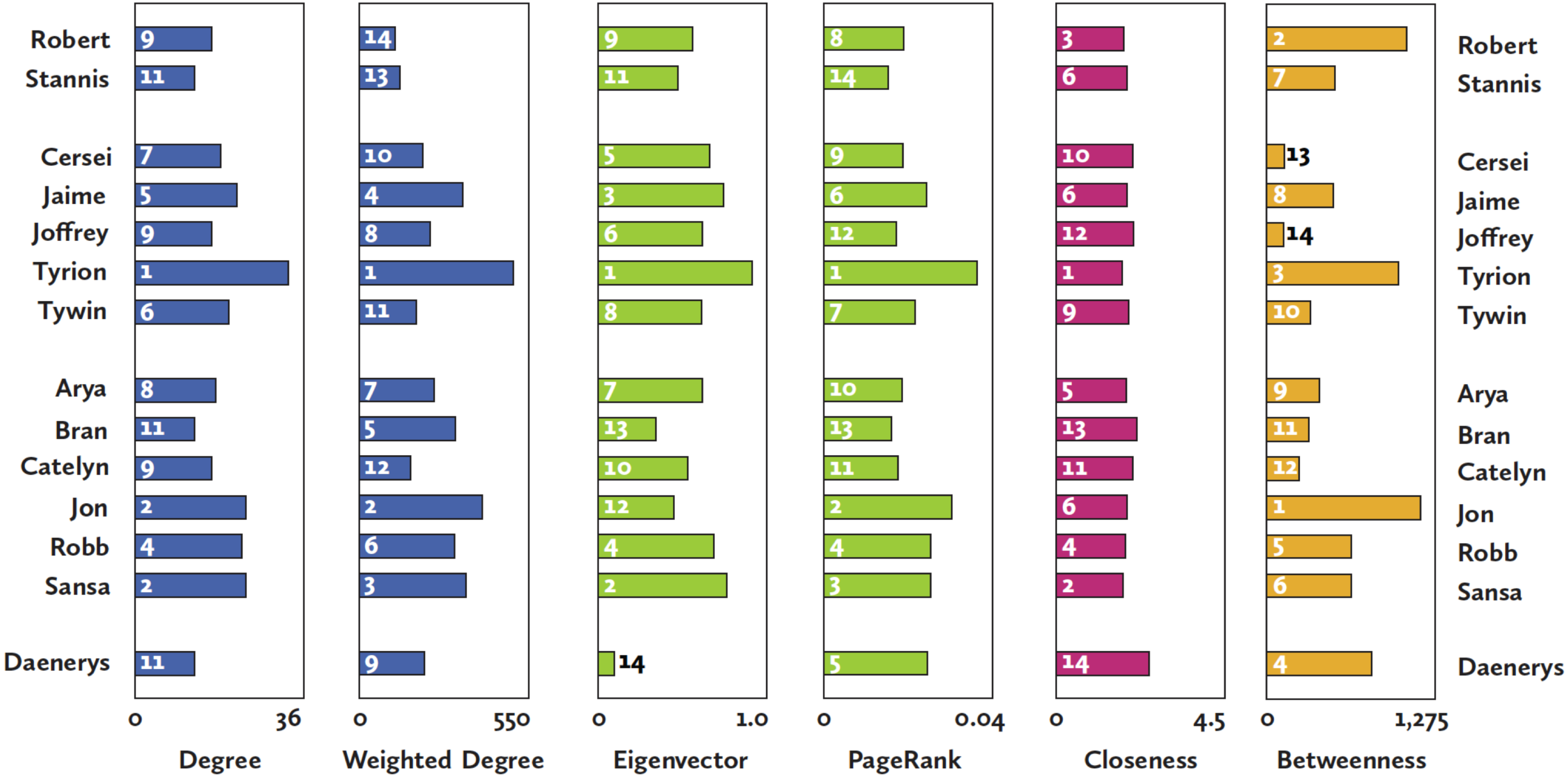
1. Jon Snow
2. Tyrion
3. Robb
4. Jaime
5. Joffrey
6. Sansa
7. Daenerys
8. Samwell
9. Arya
10. Catelyn



DATA FROM ANDREW BEVERIDGE, HTTPS://GITHUB.COM/MATHBEVERIDGE/ASOIAF

GAME OF THRONES – A STORM OF SWORDS

BEVERIDGE & SHAN LOOKED AT A SUBSET OF CHARACTERS

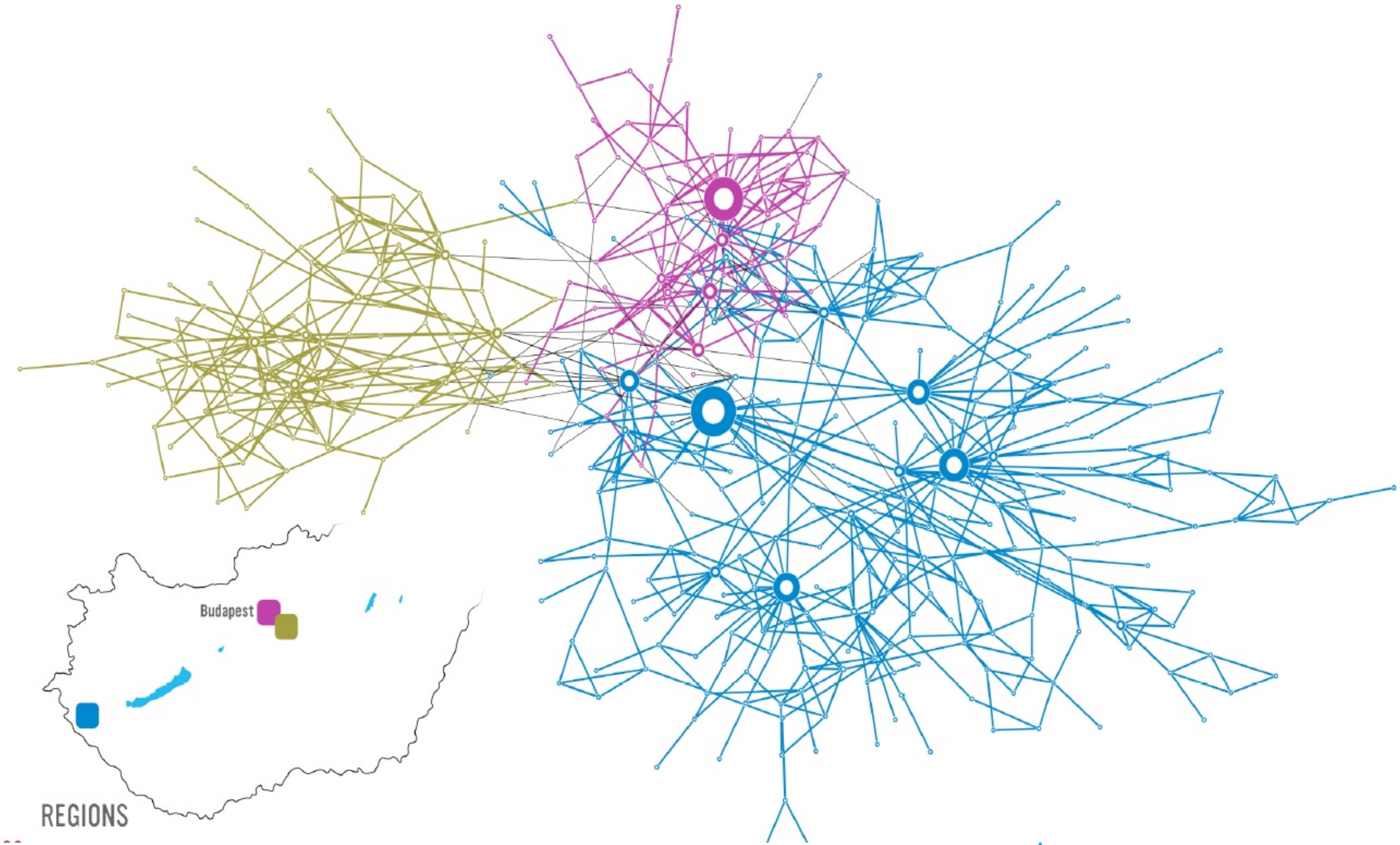


A. BEVERIDGE, J. SHAN. NETWORK OF THRONES. *MATH HORIZONS*, **APRIL** (2016) 18-22.

EXERCISE

For each of the following, come up with an example network and a particular node in that network that has:

1. High closeness centrality, but low degree centrality.
2. High degree centrality, but low closeness centrality.
3. High betweenness centrality, but low closeness centrality.
4. High closeness centrality, but low betweenness centrality.
5. High degree centrality, but low betweenness centrality.
6. High betweenness centrality, but low degree centrality.



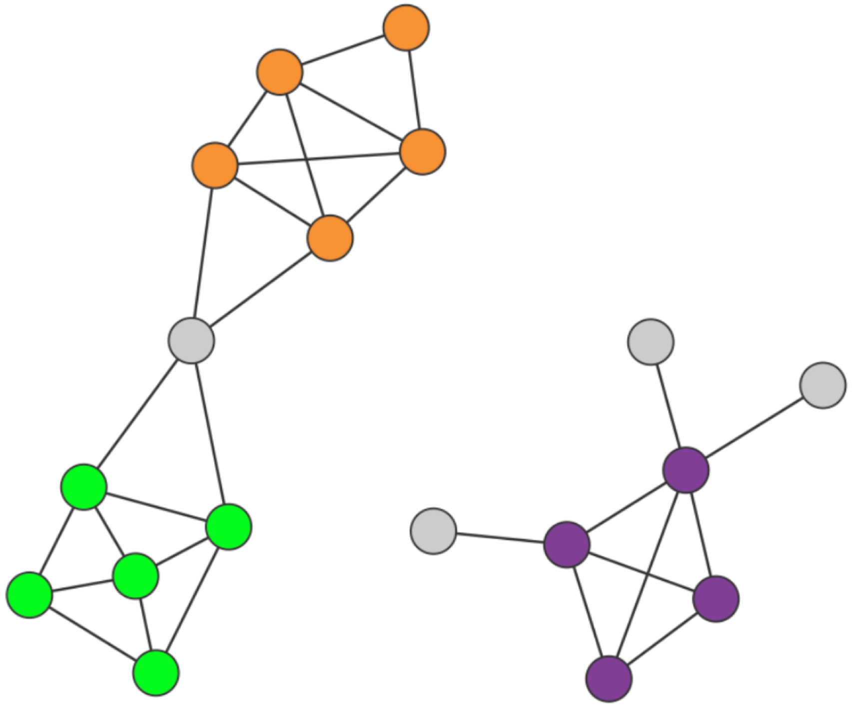
COMMUNITIES

Network Science
Summer Research Institute 2019

IMAGE FROM ALBERT-LASZLO BARABASI'S *NETWORK SCIENCE.* (2016) CAMBRIDGE UNIVERSITY PRESS.

HYPOTHESES FOR COMMUNITIES

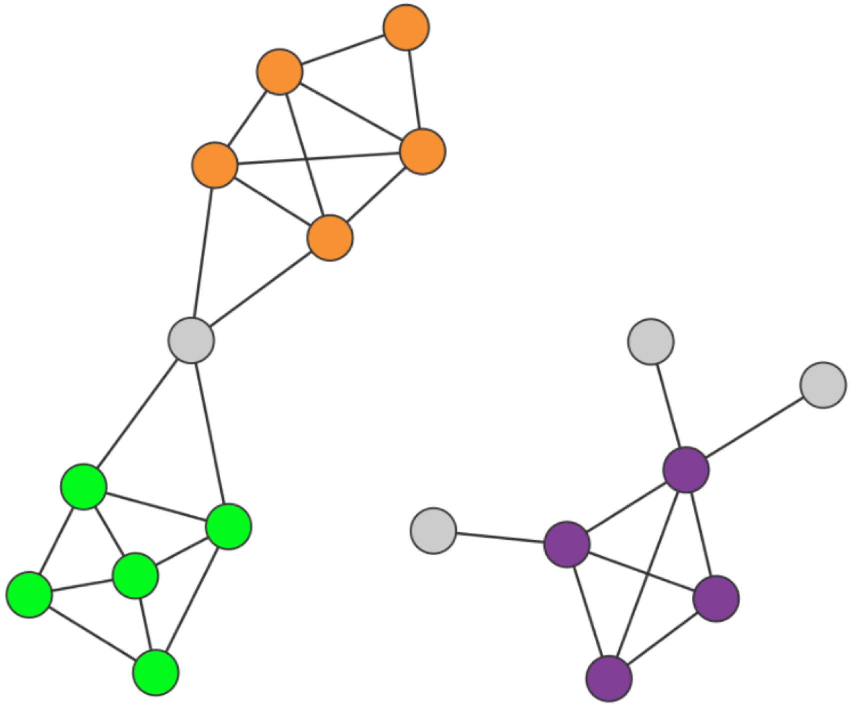
1. A network's community structure is uniquely encoded in its wiring diagram.
2. A community corresponds to a connected subgraph.
   - All members of a community are connected by a path that stays in the community.
3. Communities correspond to locally dense neighborhoods of a network.
   - Nodes in a community have a higher probability of linking to each other than to nodes not in the community.



MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

CLIQUES AS COMMUNITIES?

- A clique is a complete subgraph of k nodes.
- Triangles are frequent; larger cliques are rare.
- Communities do not necessarily correspond to complete subgraphs.



MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

STRONG AND WEAK COMMUNITIES

Consider a connected subgraph, C, with N_C nodes.

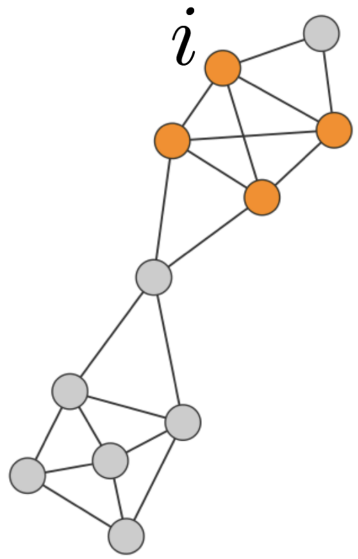
Internal Degree, k_i^{int}, is the number of links incident with node *i*, that connect to other nodes in C.

External Degree, k_i^{ext}, is the number of links incident with node *i*, that connect to nodes not in C.

If $k_i^{ext} = 0$, then all neighbors of *i* belong to C and C is a good community for *i*.

If $k_i^{int} = 0$, then all neighbors of *i* belong to other communities and C is not a good community for *i*.

$$k_i^{int} = 3$$
$$k_i^{ext} = 1$$



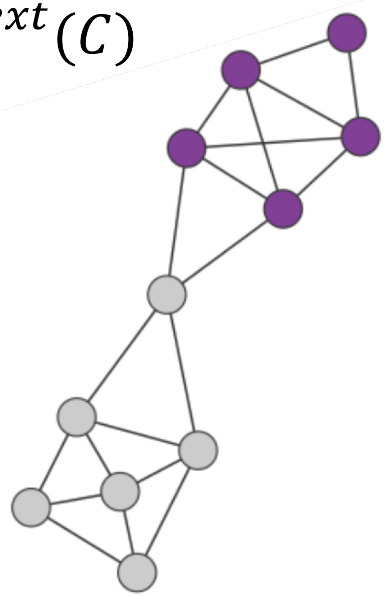
MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

STRONG AND WEAK COMMUNITIES

Strong Community: Each node of C has more links within the community than with the rest of the graph

$k_i^{int}(C) > k_i^{ext}(C)$

For all $i \in C$.

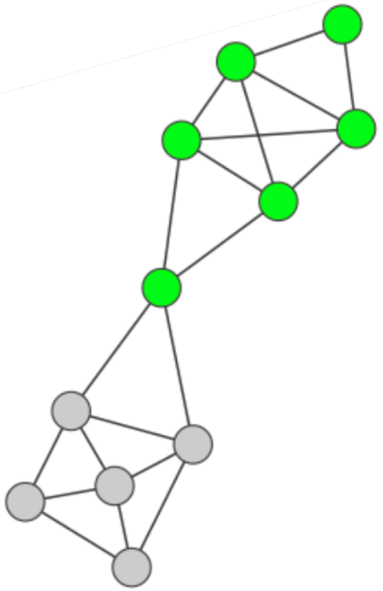


Weak Community: The total internal degree of C is greater than the total external degree.

$\sum_{i \in C} k_i^{int}(C) > \sum_{i \in C} k_i^{ext}(C)$

This is a relaxation of the strong community. It allows some vertices to violate $k_i^{int}(C) > k_i^{ext}(C)$

Every strong community is a weak community.



MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

FIRST IDEA TO FIND COMMUNITIES – GRAPH PARTITIONING

How many ways are there to partition a network into two communities?

<u>Graph Bisection</u>:

Divide a network into two equal non-overlapping subgraphs such that the number of links between the nodes in the two groups is minimized.

Two subgroups of sizes n_1 and n_2, total number of combinations $\frac{N!}{n_1!n_2!}$.

When $n_1 = n_2 = N/2$, this is approximately $\frac{2^{N+1}}{\sqrt{N}}$.

When N=10, this would give 256 partitions (1 ms).

When N=100, this would give 10^{26} partitions (10^{21} years).

FIRST IDEA TO FIND COMMUNITIES – GRAPH PARTITIONING

Community Detection

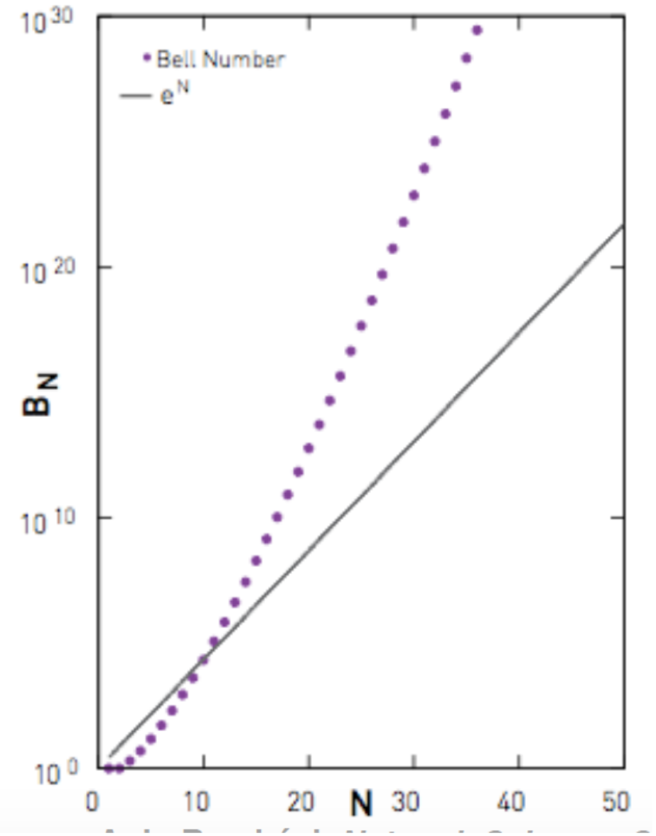
The number and size of communities are unknown at the beginning.

Partition

Division of a network into groups of nodes, so that each node belongs to to one group.

Bell Number: number of possible partitions of N nodes

$$B_N = \frac{1}{e}\sum_{j=0}^{N}\frac{j^N}{j!}$$



MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

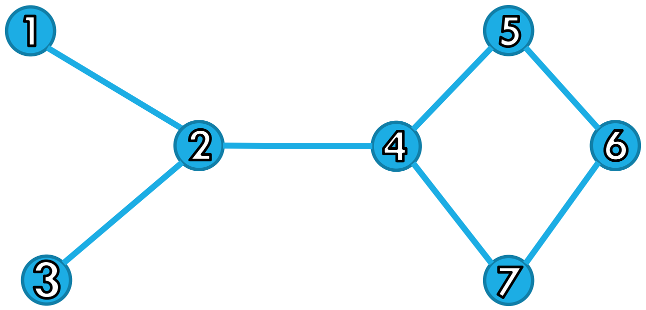
SECOND IDEA TO FIND COMMUNITIES – HIERARCHICAL CLUSTERING

1. Determine how similar nodes are using the adjacency matrix.
2. Hierarchical clustering iteratively identifies groups of nodes with high similarity, following one of two strategies:
   a) Agglomerative Algorithms: Merge nodes and communities with high similarity.
   b) Divisive Algorithms: Split communities by removing links that connect nodes with low similarity.
3. A hierarchical tree or dendrogram is used to visualize the history of the merging or splitting process the algorithm follows. Horizontal cuts of this tree offer various community partitions.

GIRVAN-NEWMAN ALGORITHM - DIVISIVE

1. Define a centrality measure for the edges.
   - Link betweenness – the number of shortest paths between all node pairs that run along a link.
2. Compute the centrality of each link. Remove the link with the largest centrality; in case of a tie, choose randomly.
3. Recalculate the centrality of each link.
4. Repeat until all links are removed.

Progress can be represented using a tree or *dendrogram*.

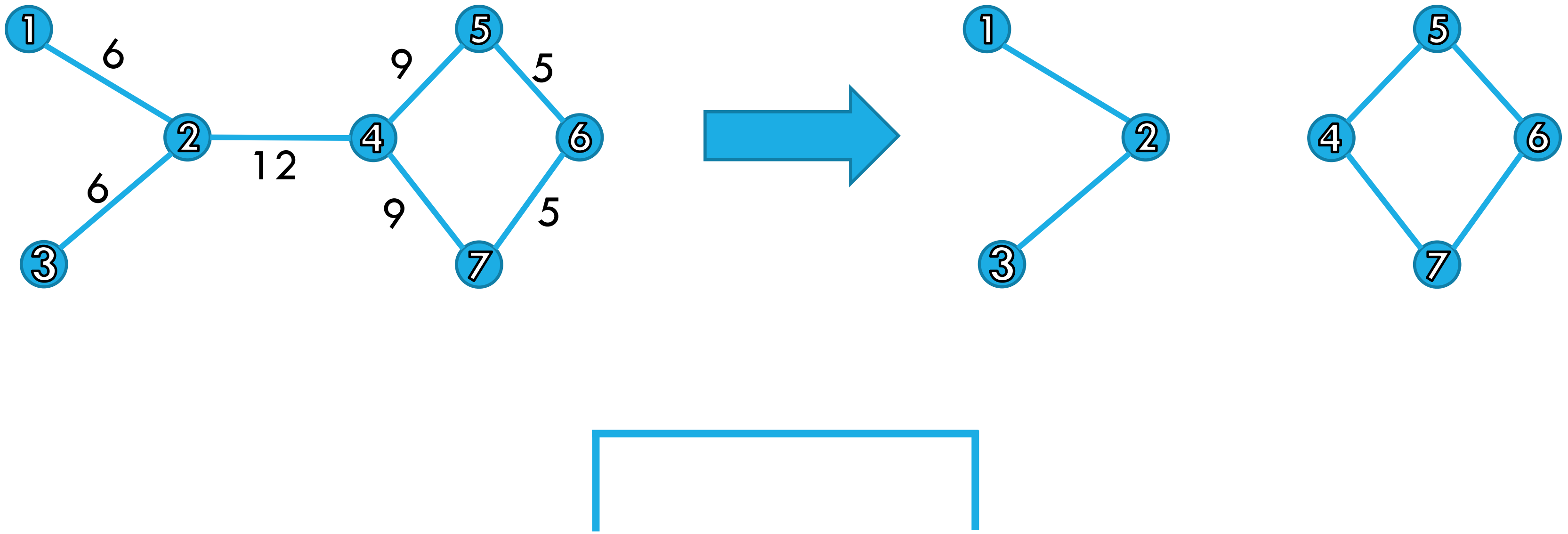


MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)
M.E.J. NEWMAN. (2010). *NETWORKS: AN INTRODUCTION.* OXFORD UNIVERSITY PRESS.

GIRVAN-NEWMAN ALGORITHM - DIVISIVE

1. Define a centrality measure for the edges.
2. Compute the centrality of each link.
   Remove the link with the largest centrality; in case of a tie, choose randomly.
3. Recalculate the centrality of each link.
4. Repeat until all links are removed.

Progress can be represented using a tree or *dendrogram*.

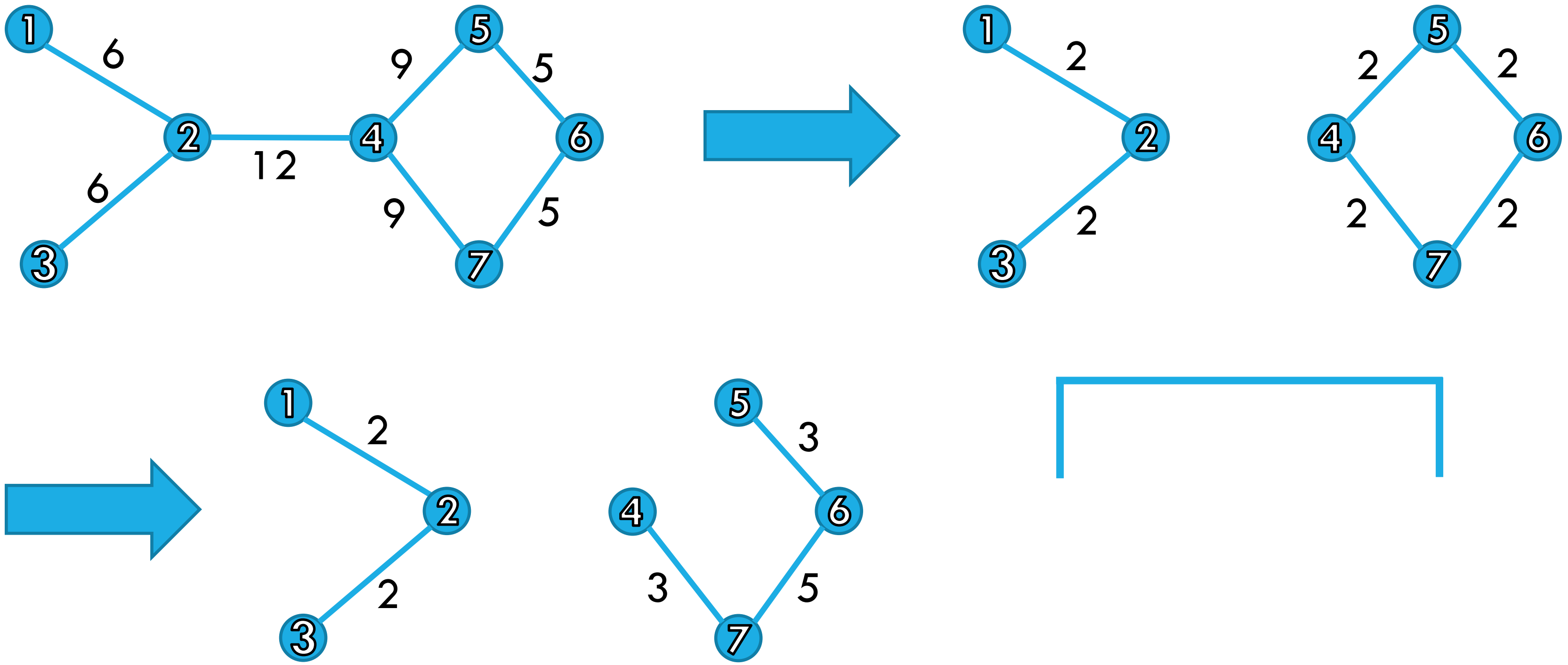


MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)
M.E.J. NEWMAN. (2010). *NETWORKS: AN INTRODUCTION.* OXFORD UNIVERSITY PRESS.

GIRVAN-NEWMAN ALGORITHM - DIVISIVE

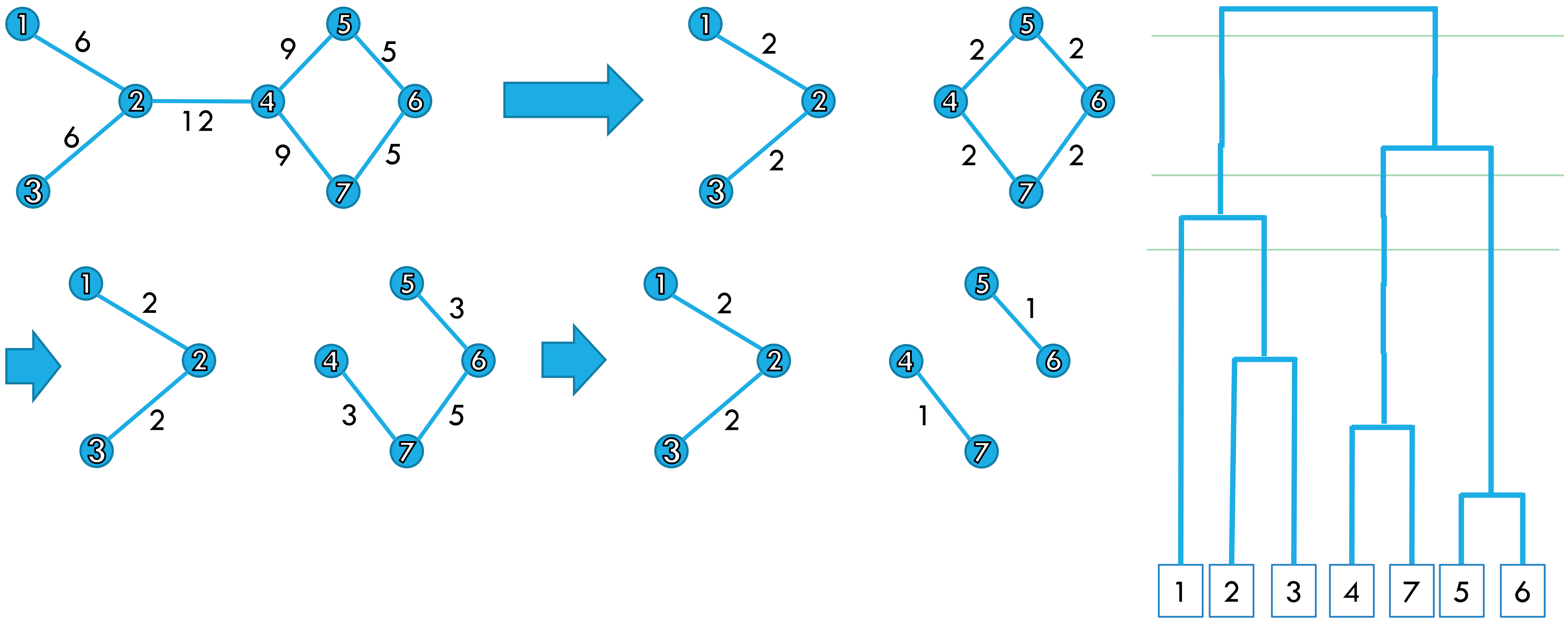
1. Define a centrality measure for the edges.
2. Compute the centrality of each link.
   Remove the link with the largest centrality; in case of a tie, choose randomly.
3. Recalculate the centrality of each link.
4. Repeat until all links are removed.

Progress can be represented using a tree or *dendrogram*.

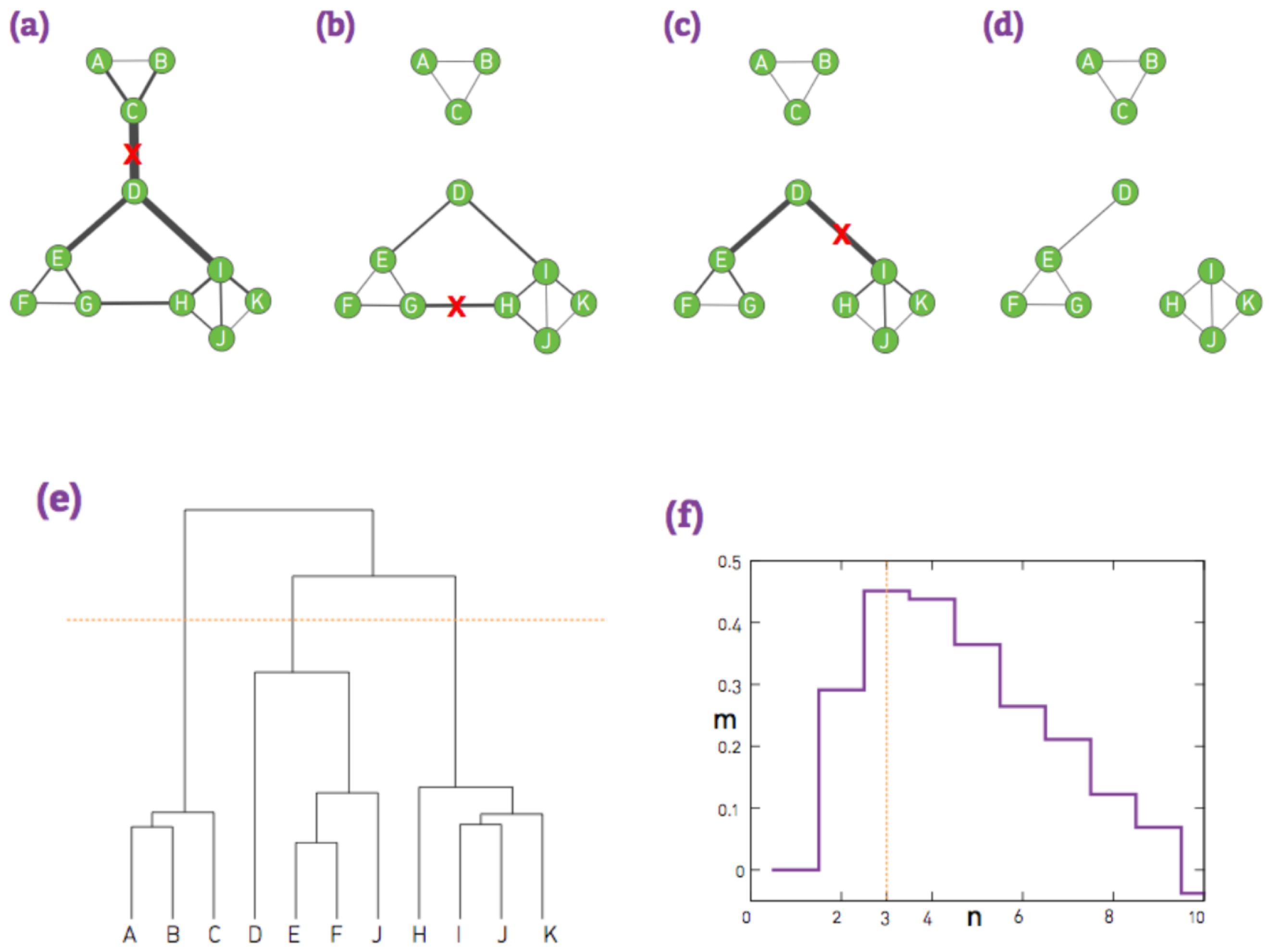


MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)
M.E.J. NEWMAN. (2010). *NETWORKS: AN INTRODUCTION.* OXFORD UNIVERSITY PRESS.

GIRVAN-NEWMAN ALGORITHM - DIVISIVE

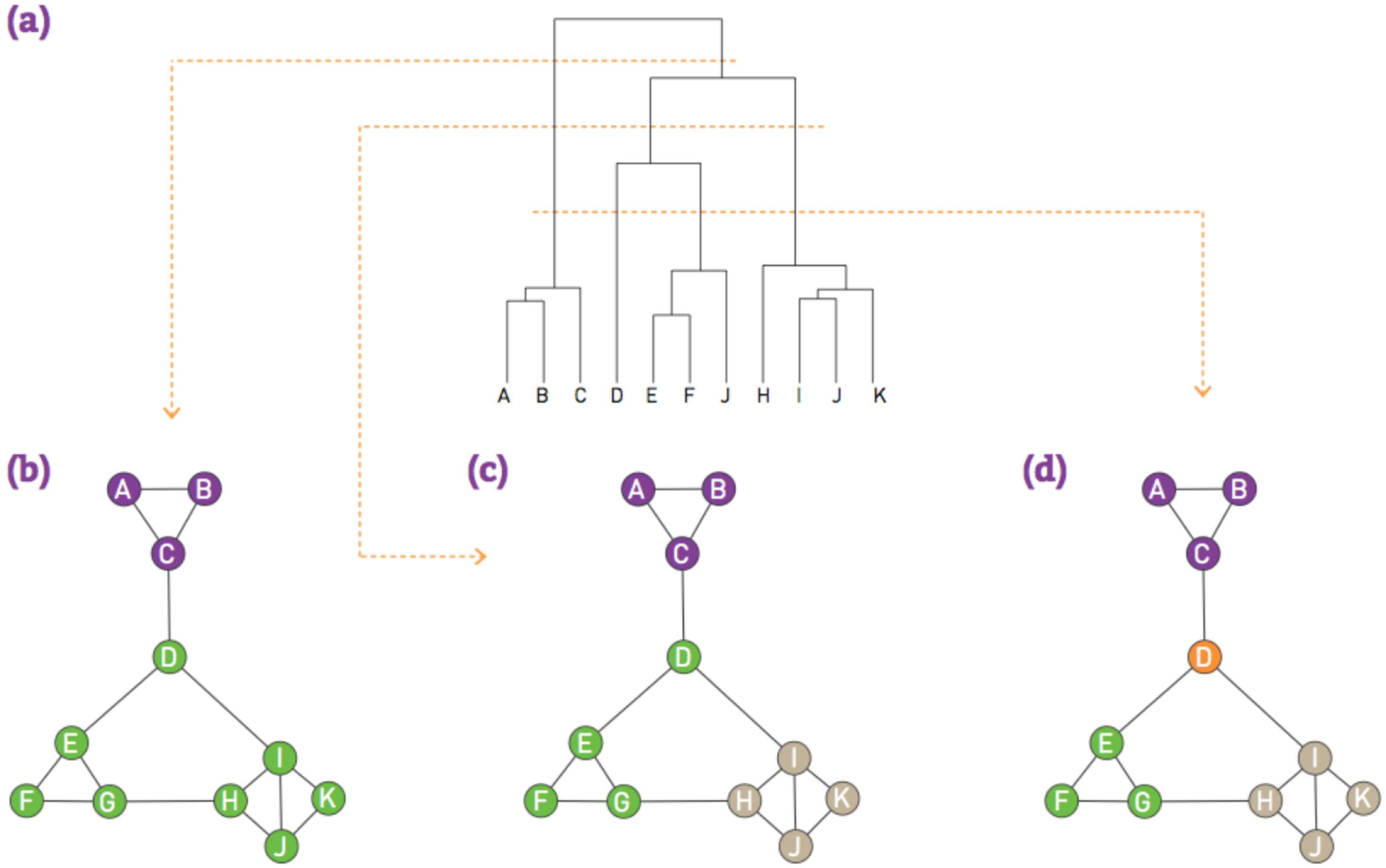


GIRVAN-NEWMAN



MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

WHERE TO CUT?



MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

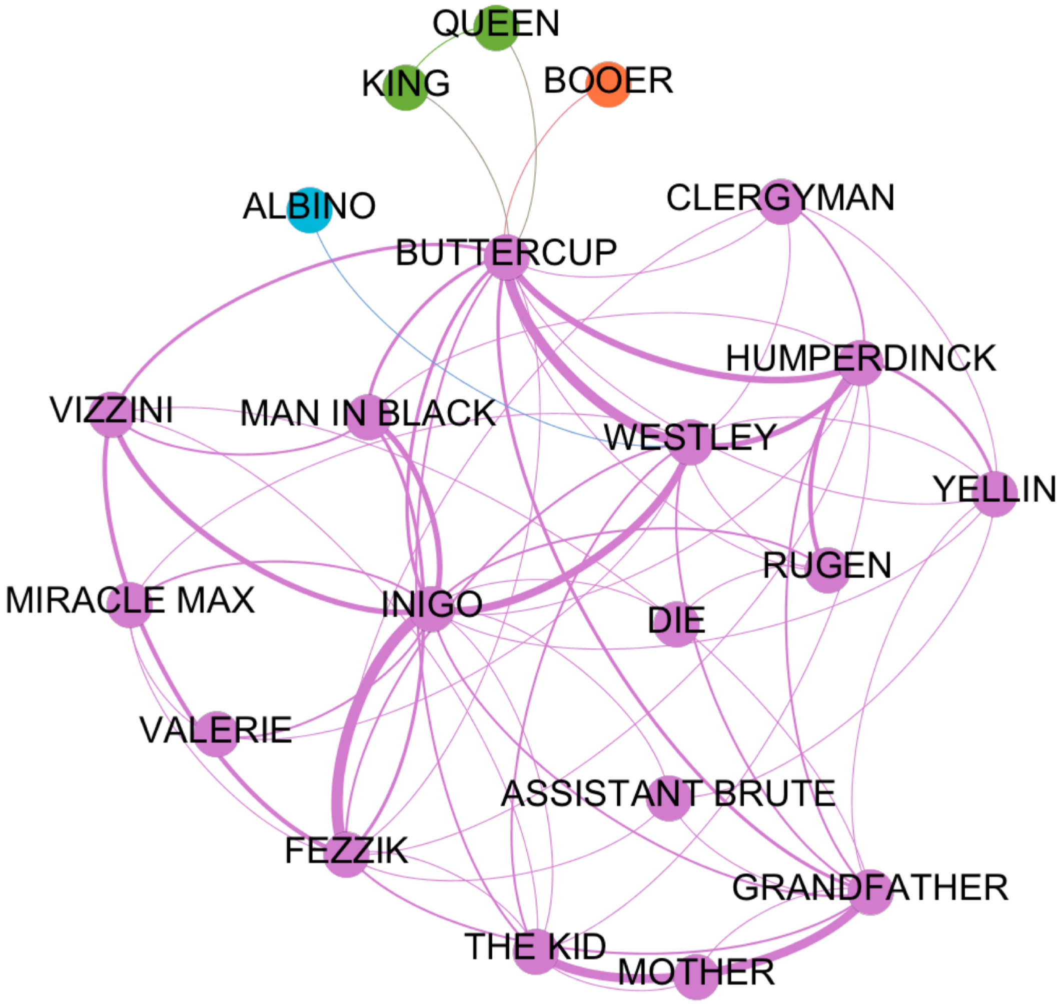
GIRVAN-NEWMAN IN GEPHI

Add the plugin Newman-Girvan Clustering

1. Open the Tools menu
2. Choose Plugins
3. Search for "Newman-Girvan" and install

Run the Girvan-Newman Algorithm

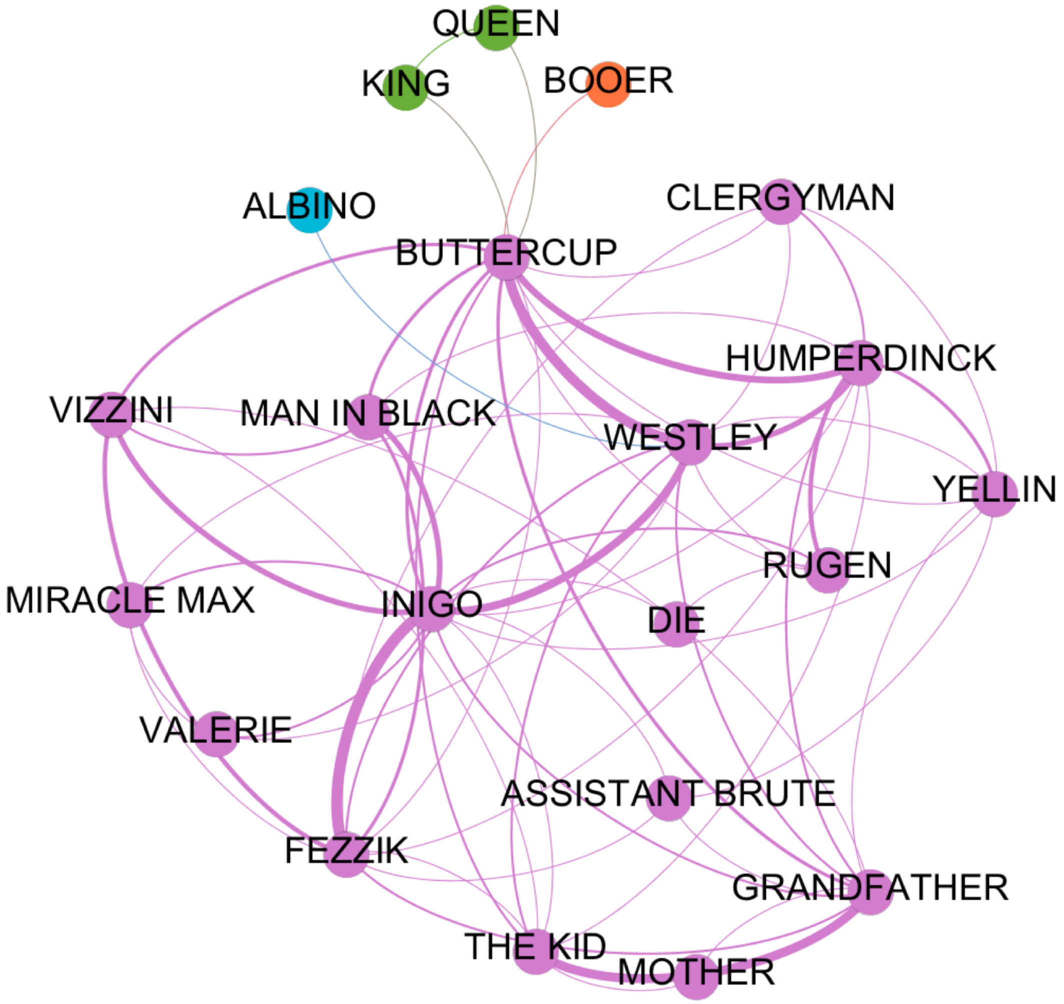
1. Under the Statistics Panel, click "Run" next to "Givan-Newman-Clustering"
2. Decide if you want to respect edge type and parallel edges
3. Click "Ok"
4. In the Data Laboratory there is now a column called "Cluter-ID." The number in this column tells you to which community the node belongs.



GIRVAN-NEWMAN IN GEPHI

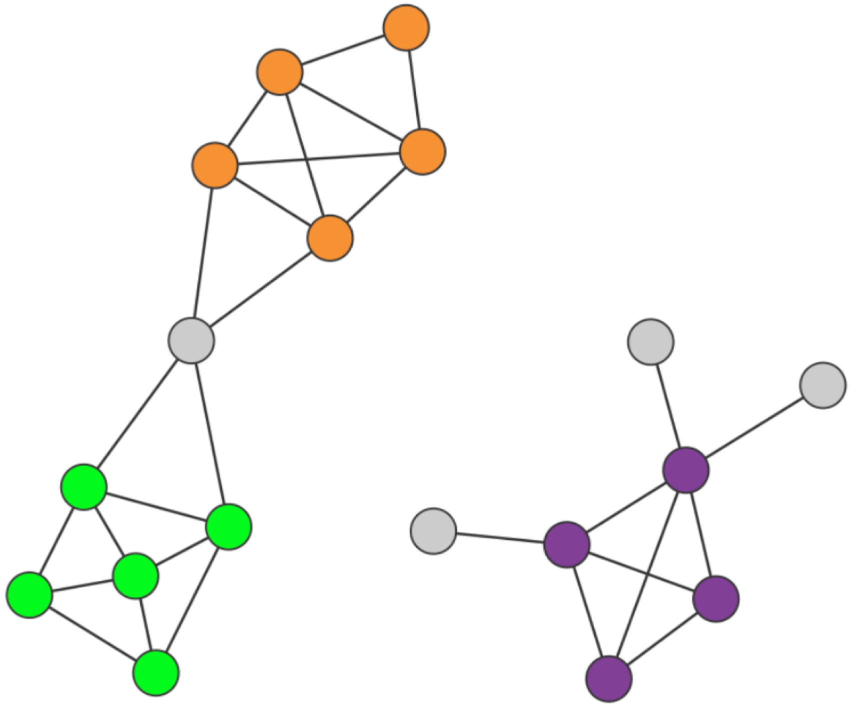
To color the nodes by the community…

1. In the Appearance Panel click the color palette.
2. Choose "Partition."
3. In the drop down menu, select "Cluter-ID."
4. Here you can also see the percentage of nodes in each community.
5. Click "palette" to change the colors or generate a new palette.
6. Click "Apply."



HYPOTHESES FOR COMMUNITIES

1. A network's community structure is uniquely encoded in its wiring diagram.
2. A community corresponds to a connected subgraph.
   - All members of a community are connected by a path that stays in the community.
3. Communities correspond to locally dense neighborhoods of a network.
   - Nodes in a community have a higher probability of linking to each other than to nodes not in the community.



MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

MODULARITY

Add a random hypothesis:

Randomly wired networks are not expected to have a community structure.

Imagine a partition into n_C communities $\{C_C, c = 1, n_C\}$

Modularity $$M(C_c) = \frac{1}{2L}\sum_{i,j=1}^{N}(A_{ij} - P_{ij})\delta(C_i - C_j)$$

Original data

Expected connections, a model

Relative to a specific partition

Modularity is a measure associated to a partition.

MODULARITY

Maximal Modularity Hypothesis

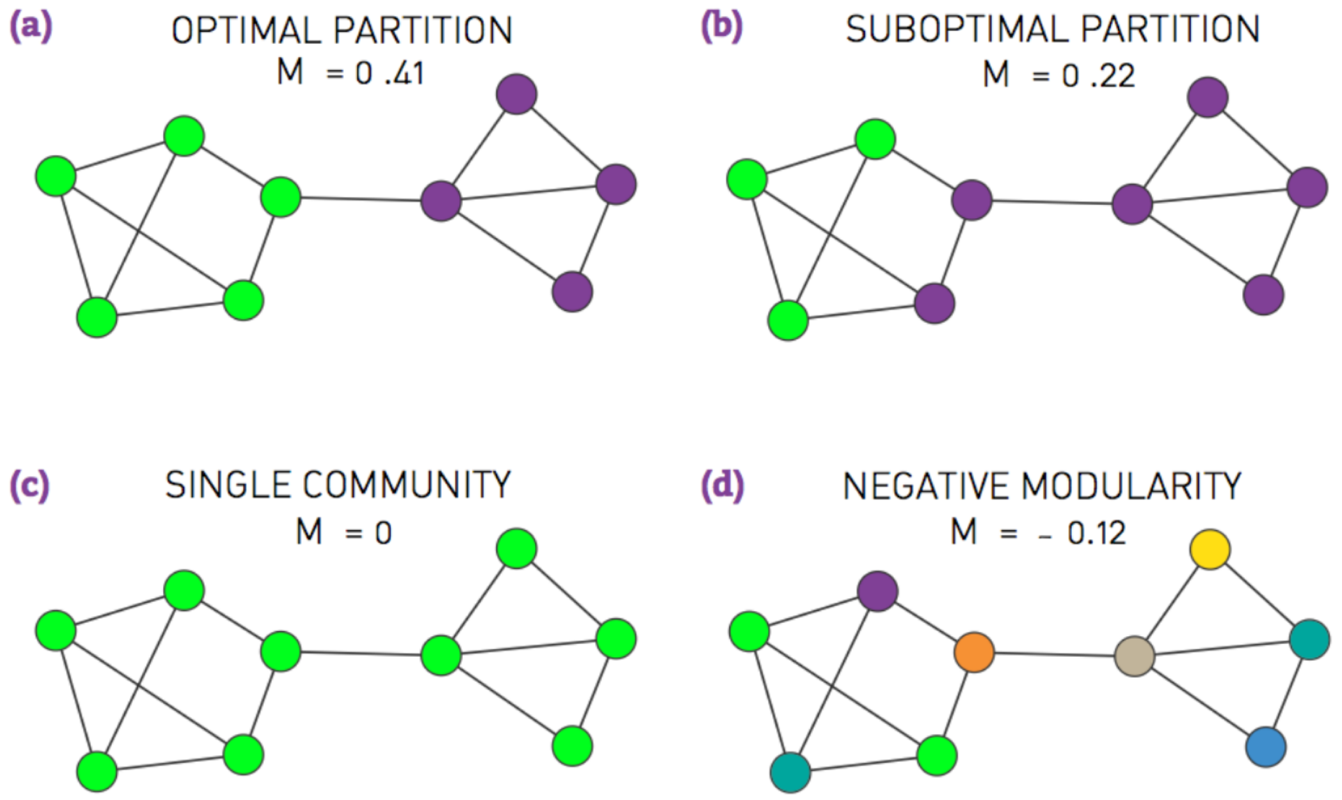
The partition with the maximum modularity M for a given network offers the optimal community structure.

Goal: Find the partition into communities that maximizes M.

MODULARITY

a) Optimal Partition – maximizes modularity
b) Suboptimal Partition – positive modularity, but not the maximum value.
c) Single Community, assigning all nodes to the same community – modularity 0
d) Assigning each node to a different community – negative modularity

Modularity is size dependent.



MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE (HTTPS://WWW.BARABASILAB.COM/COURSE)

MODULARITY BASED COMMUNITY DETECTION

<u>Greedy Algorithm</u> – iteratively join nodes if the move increases the new partitions modularity.

1. Assign each node to a community of its own. That is, start with N communities.
2. Inspect each pair of communities connected by at least one link and compute the modularity variation, ΔM, obtained if we merge these two communities.
3. Identify the community pairs for which ΔM is the largest and merge them. Modularity of a particular partition is always calculated from the full topology of the network.
4. Repeat step 2 until all nodes are merged into a single community.
5. Record for each step and select the partition for which the modularity is maximal.

There are other algorithms that are better.

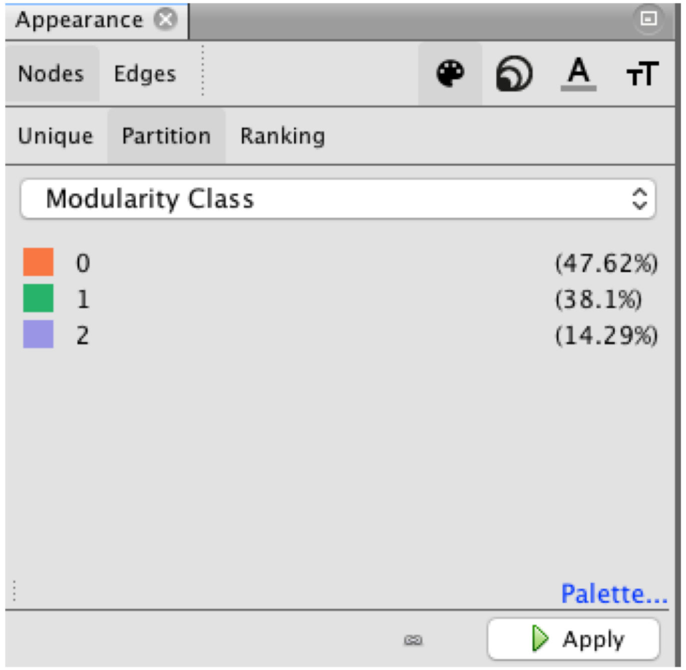
GEPHI & MODULARITY

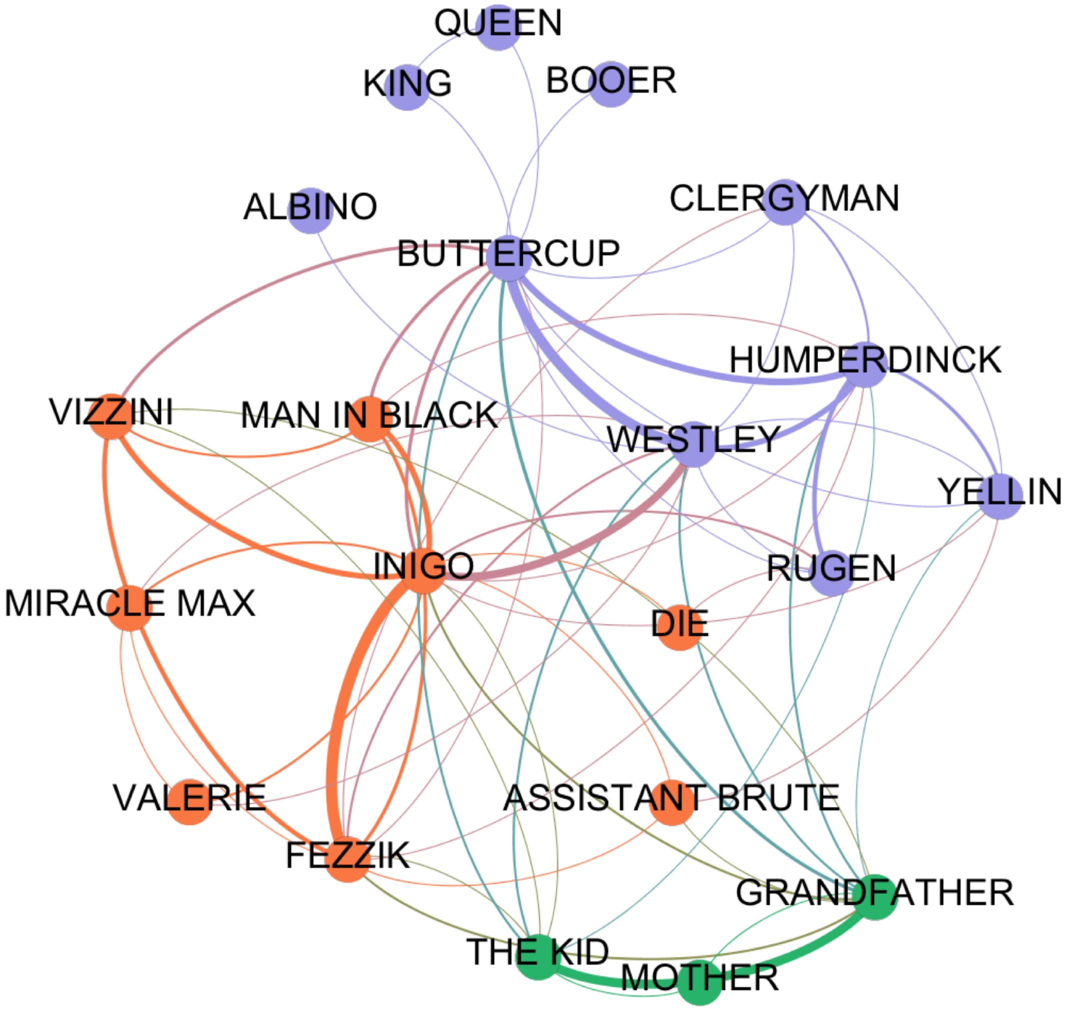
1. Under the Statistics panel click “Run” next to Modularity.
   - Choose “randomize,” choose whether or not to include edge weights. Adjust the resolution as desired.
2. Click on the “Data Laboratory” button.
   - Here you can see the modularity class of each character.
3. Under “Appearance” you can partition the nodes by their modularity class.

MODULARITY

Three communities are detected.

- Used weights
- Resolution 1

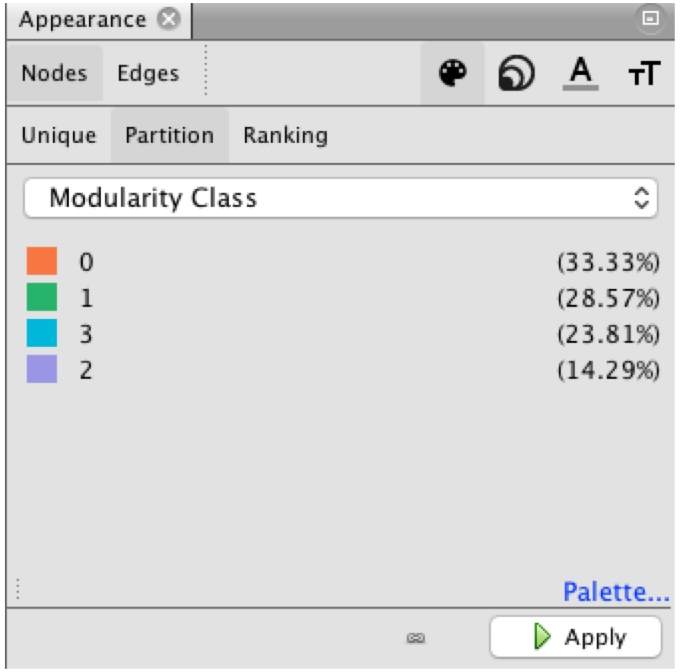


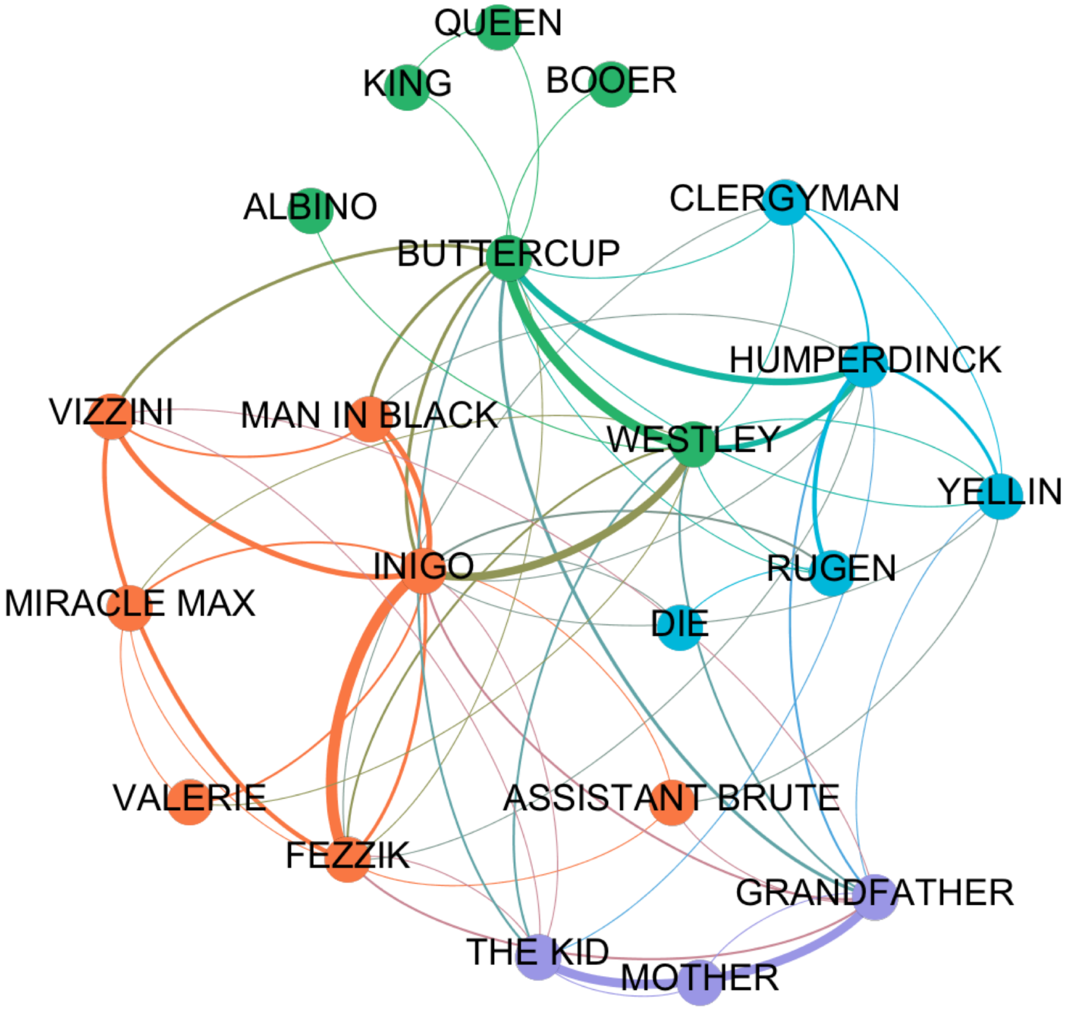


MODULARITY

Four communities are detected.

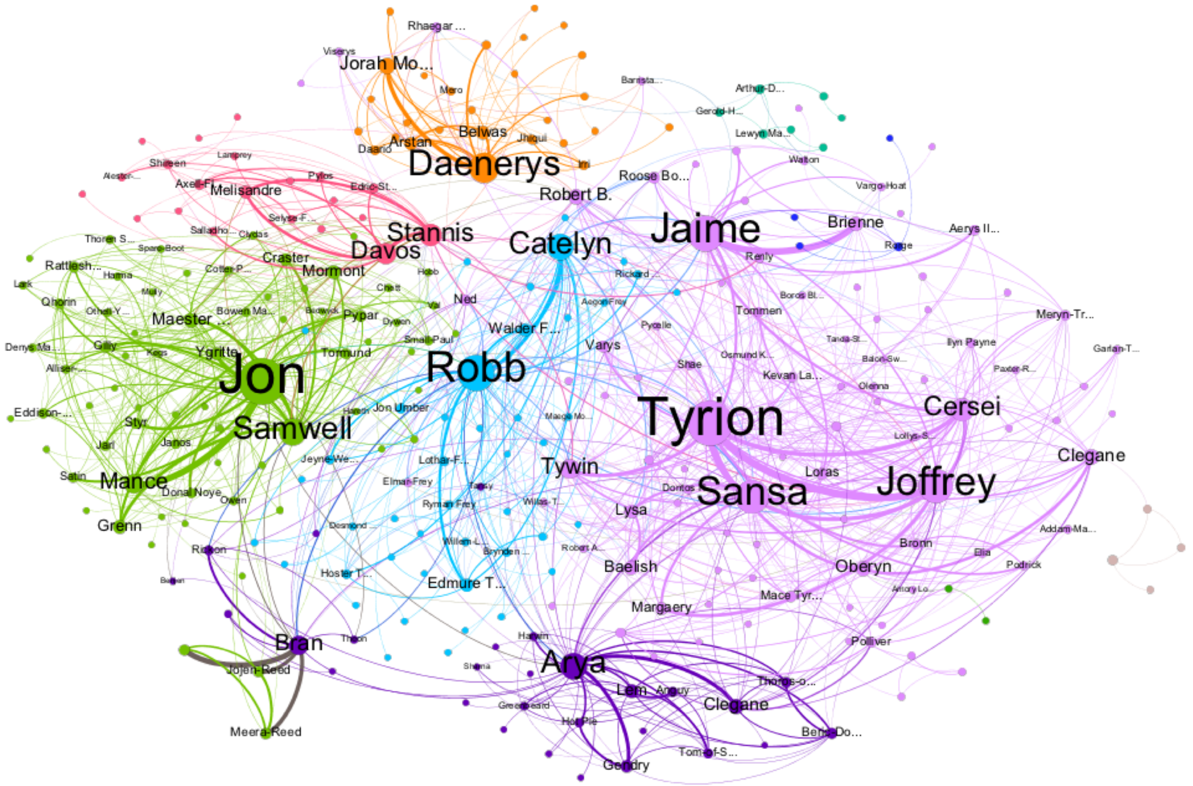
- Used weights
- Resolution 0.75



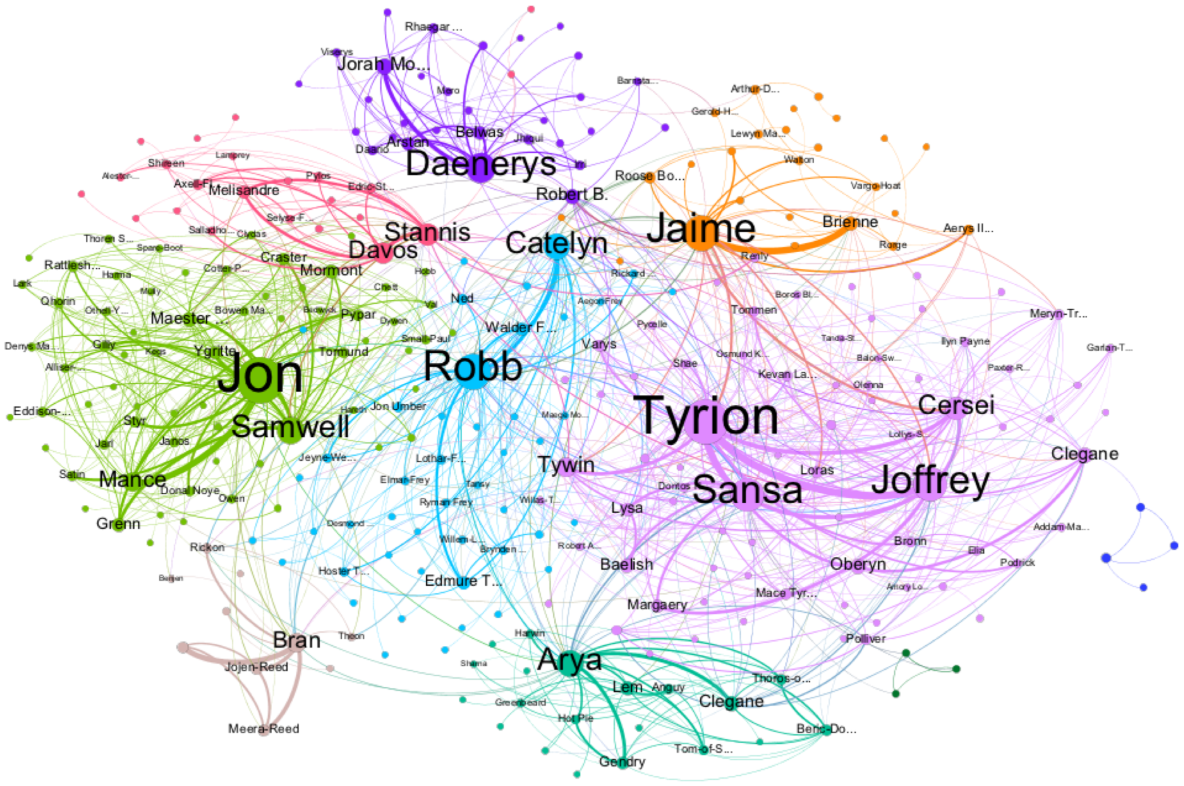


GAME OF THRONES

Girvan-Newman

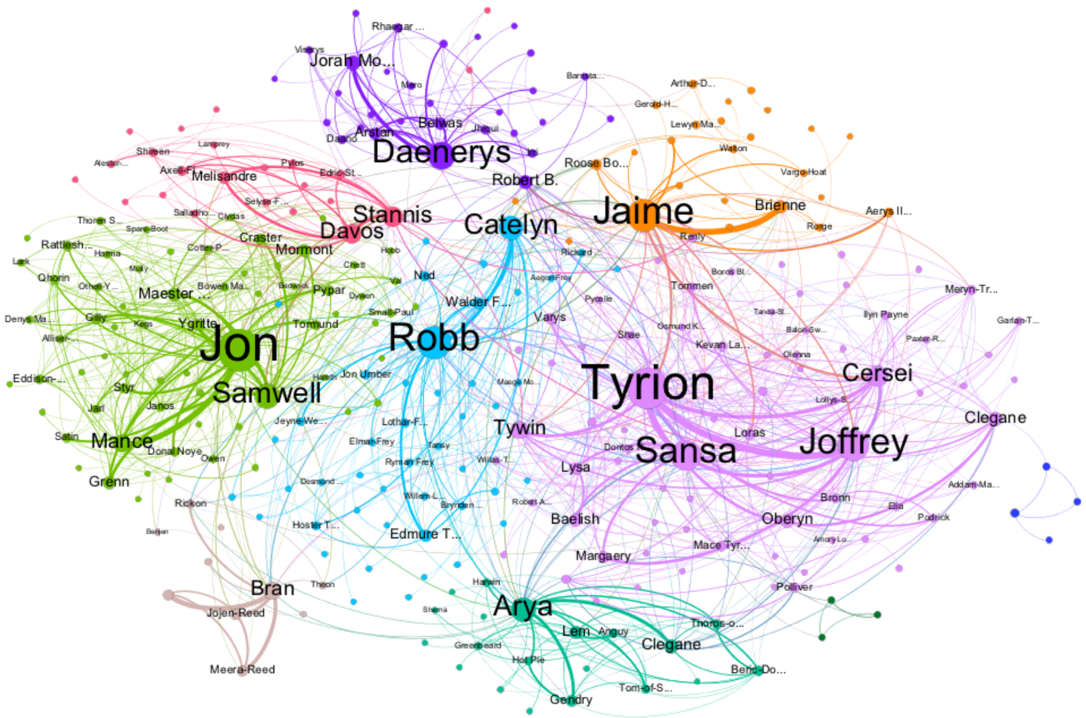


Modularity (Resolution 1, use weights)

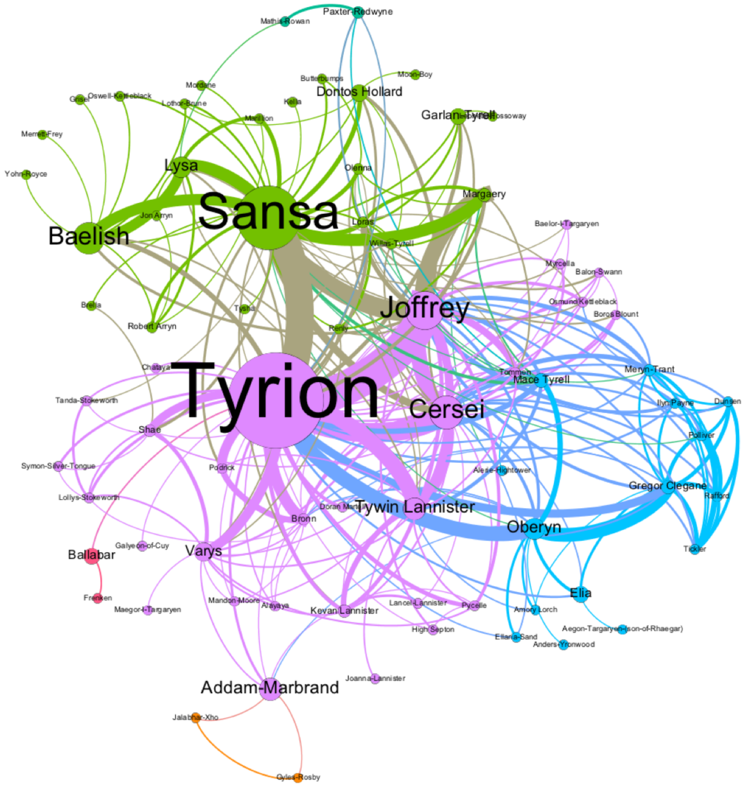


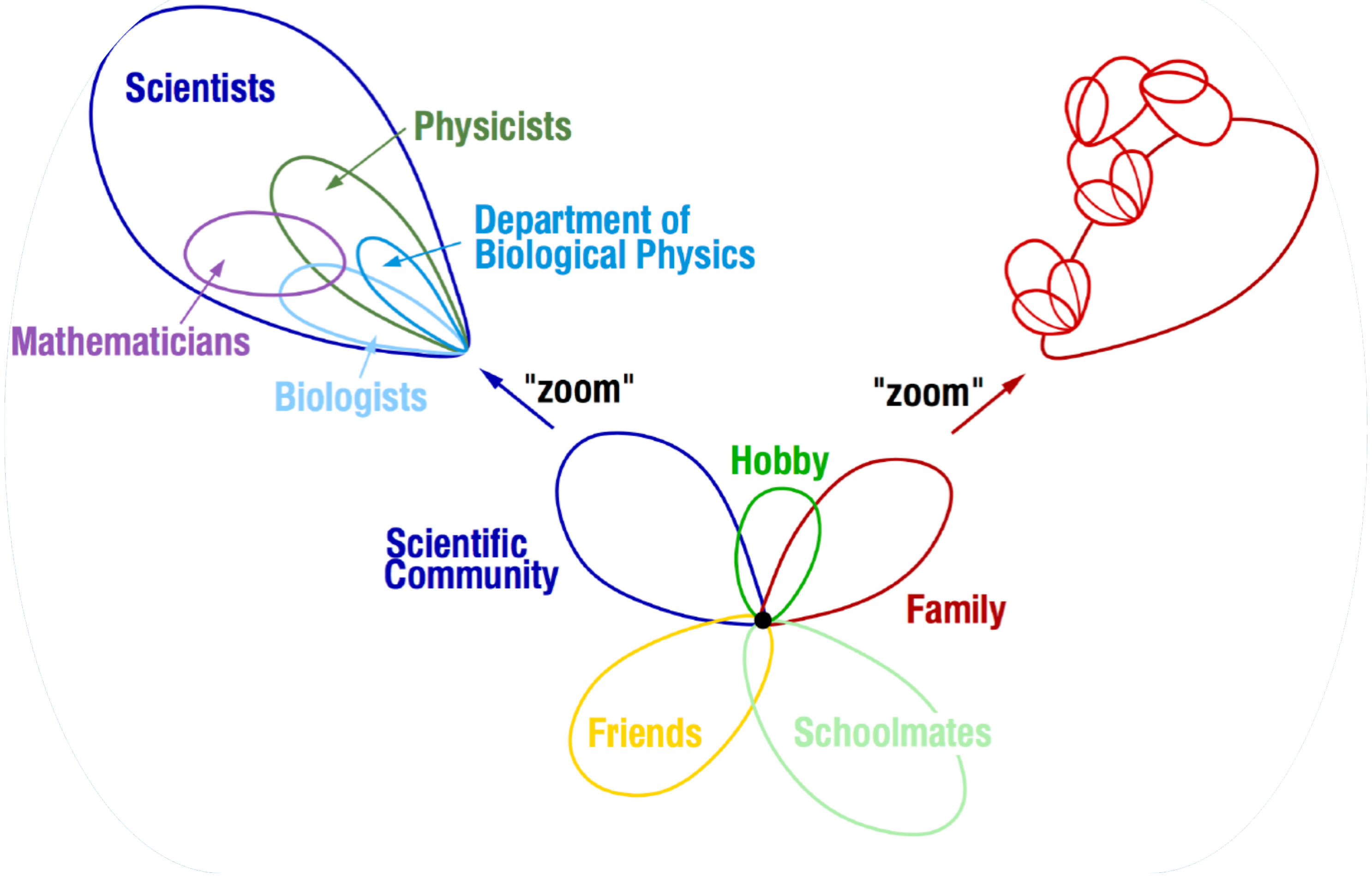
GAME OF THRONES

Modularity (Resolution 1, use weights)



Modularity Kings Landing





OVERLAPPING COMMUNITIES

Schematic representation of the communities surrounding T. Vicsek who introduced the concept of overlapping communities.

MODIFIED FROM SLIDES FOR ALBERT-LASZLO BARABASI'S COMPLEX NETWORKS COURSE ([HTTPS://WWW.BARABASILAB.COM/COURSE](https://www.barabasilab.com/course))

G. PALLA ET AL., *NATURE* 435 (2005)

CLIQUE PERCOLATION (CFINDER)

Views a community as the union of overlapping cliques.

The Cfinder package that implements Clique Percolation Method can be downloaded at:

www.cfinder.org