

PEOPLE AND NETWORKS

Research Questions

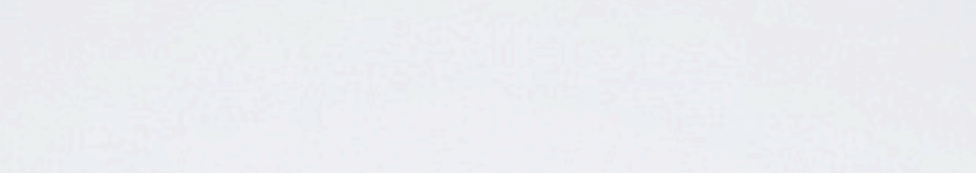
COST workshop, Oxford, 22-25 March 2015

Ruth Ahnert (QMUL)

Sebastian Ahnert (Cambridge)



NETWORKS SURROUND US



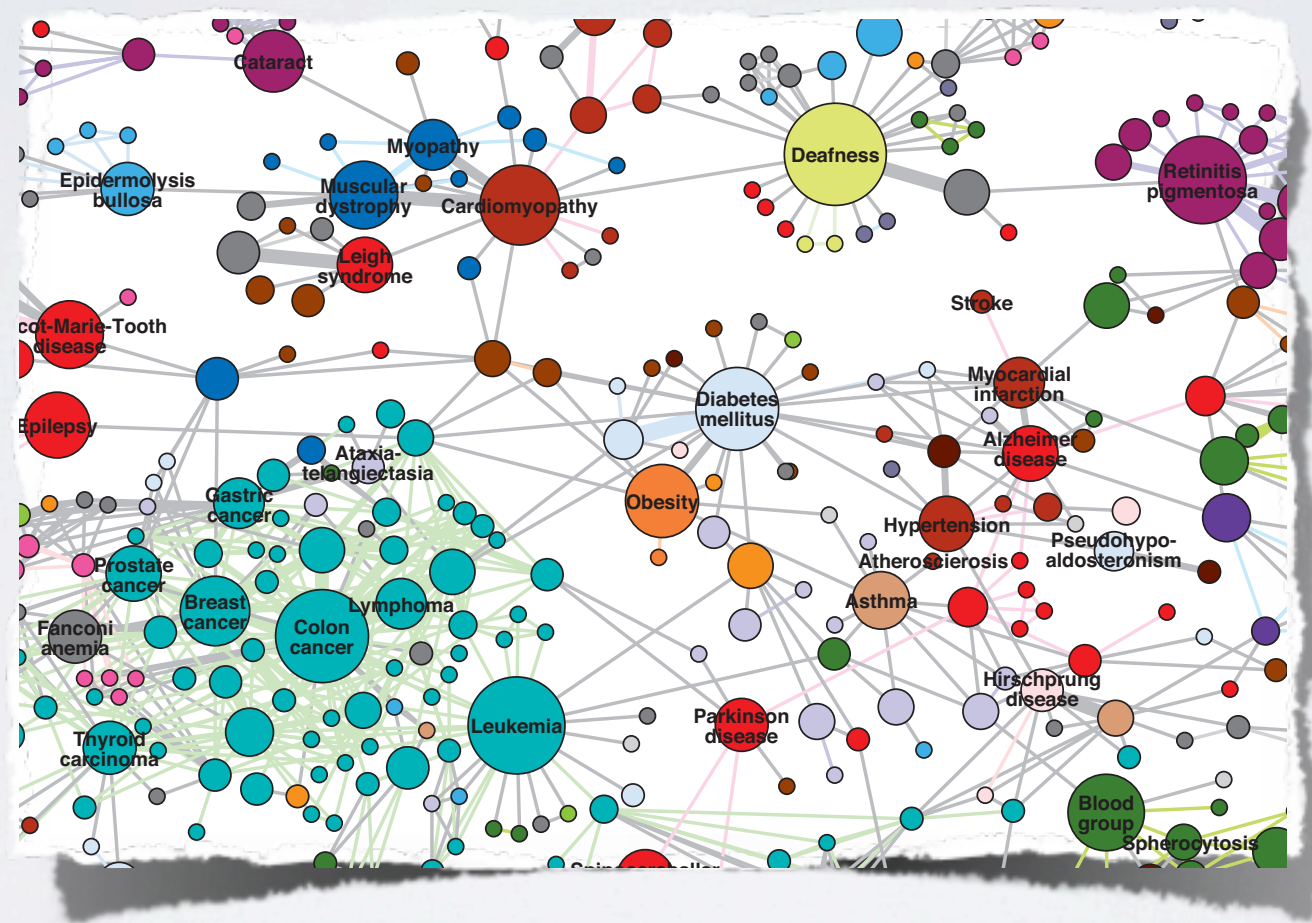
NETWORKS SURROUND US

An abundance of network data now surrounds us:

- Mobile phone networks
- Transport networks
- Power grids
- Online social networks (e.g. Facebook, Twitter)
- Gene regulatory networks
- Protein interaction networks
- Neural networks

NETWORKS SURROUND US

An abundance of network data now surrounds us:

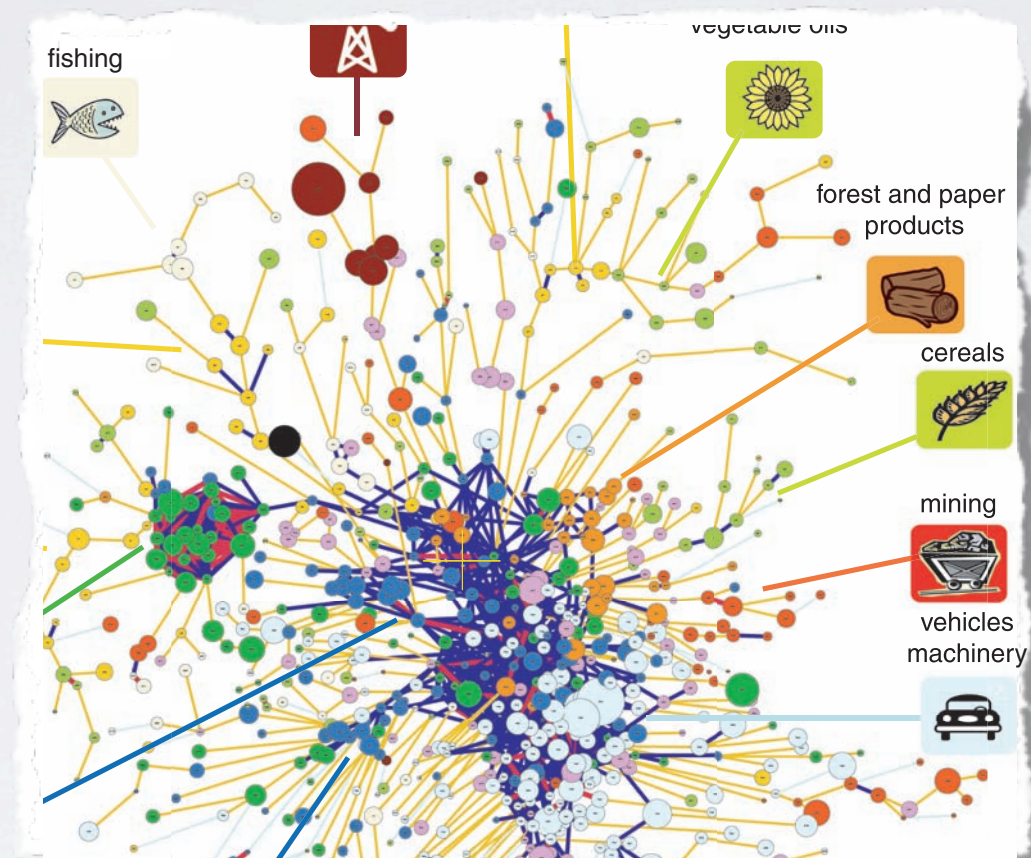


Diseases

Goh, K.-I. et al. The human disease network.
P Natl Acad Sci Usa 104, 8685–8690 (2007).

NETWORKS SURROUND US

An abundance of network data now surrounds us:

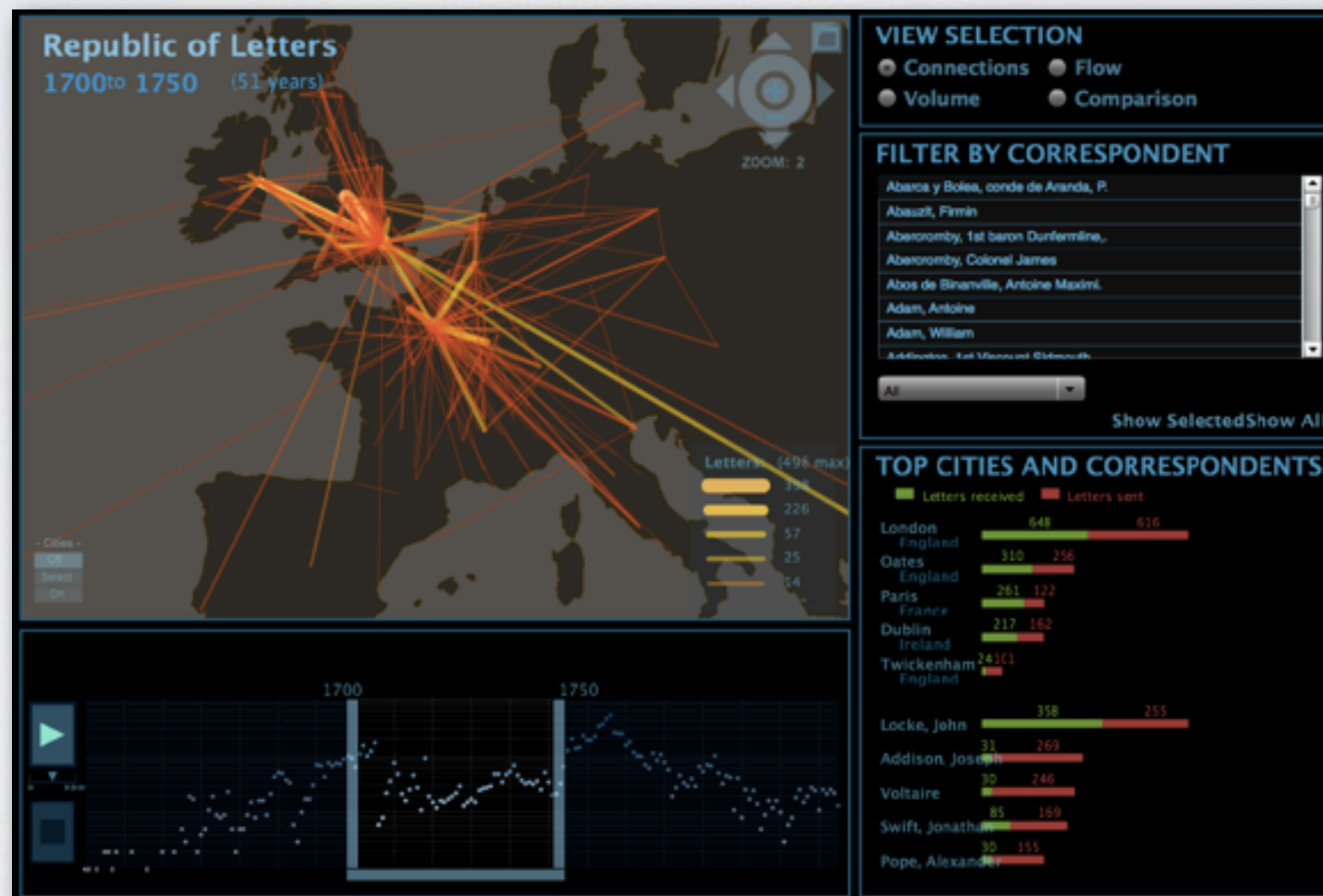


Products

Hidalgo, C. A., Klinger, B., Barabasi, A.-L. & Hausmann, R.
The product space conditions the development of nations. *Science* 317, 482–487 (2007).

EARLY MODERN NETWORKS

COST brings together people interested in particular kinds of networks:



NETWORKS

Networks

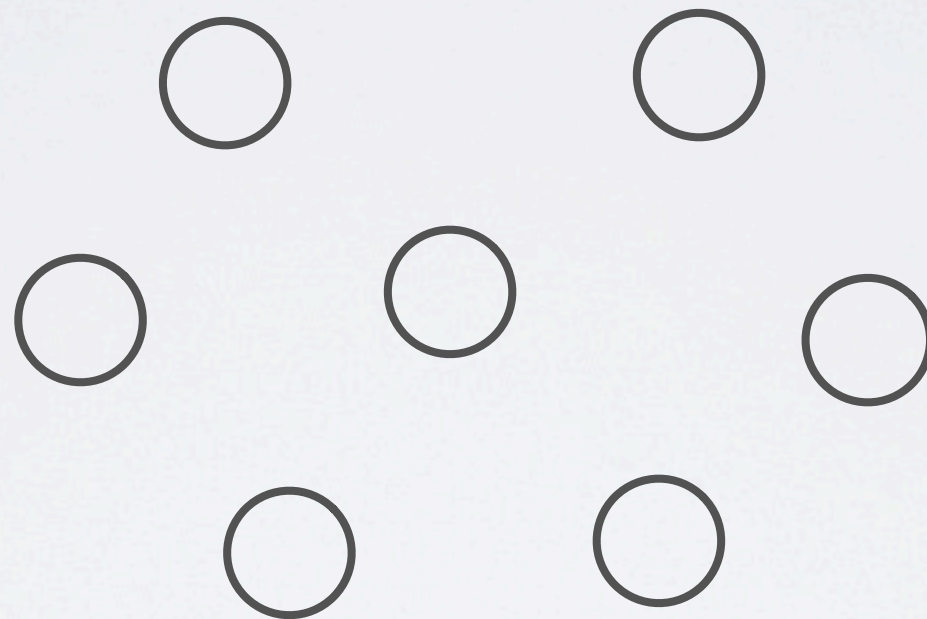
NETWORKS

Network analysis is a highly interdisciplinary research field in its own right.

NETWORKS

Network analysis is a highly interdisciplinary research field in its own right.

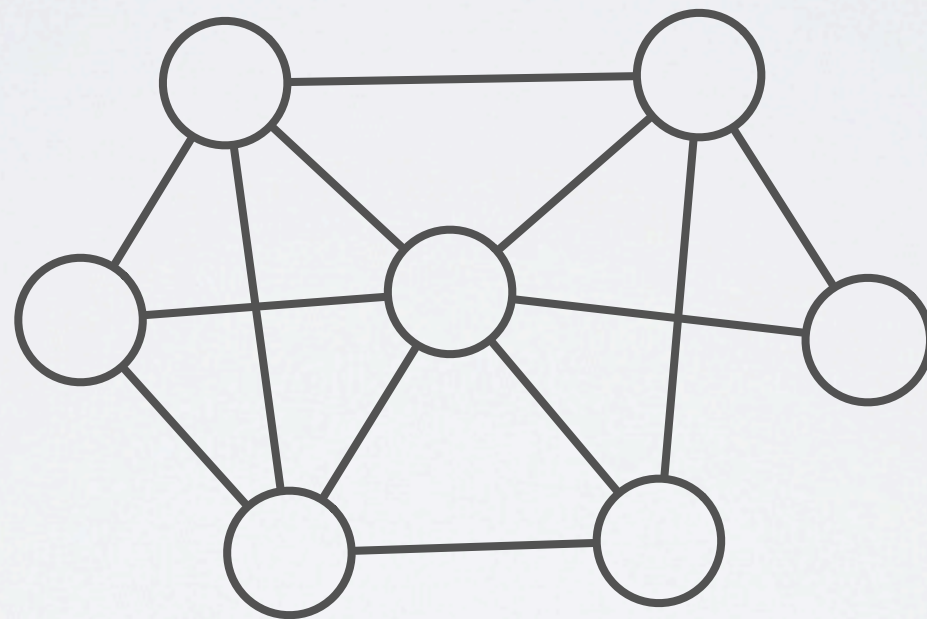
Networks consist of **nodes**



NETWORKS

Network analysis is a highly interdisciplinary research field in its own right.

Networks consist of **nodes** and **edges**.

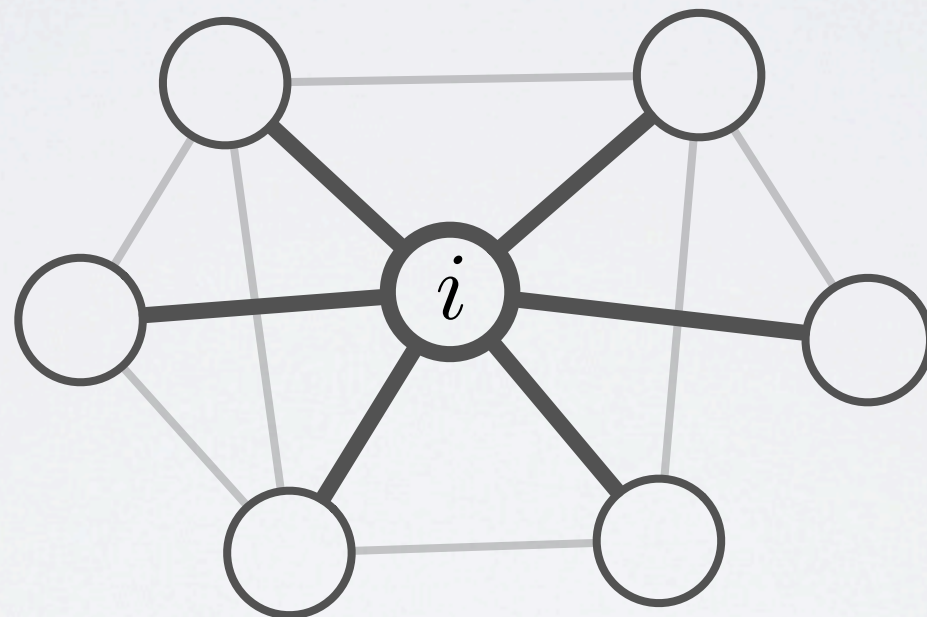


NETWORKS

Network analysis is a highly interdisciplinary research field in its own right.

Networks consist of **nodes** and **edges**.

$$k_i = 6$$



The number of connections a node has is its **degree** k .

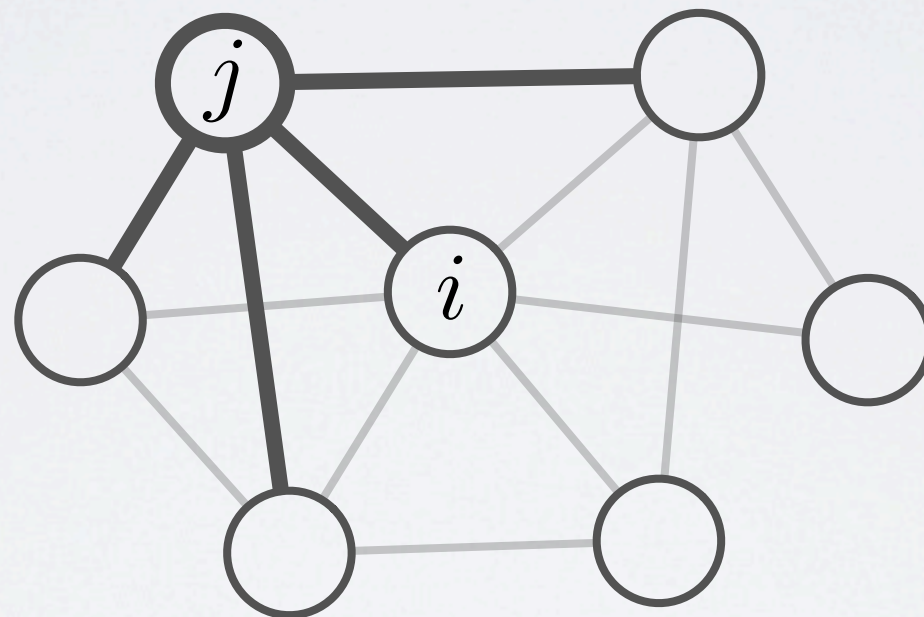
NETWORKS

Network analysis is a highly interdisciplinary research field in its own right.

Networks consist of **nodes** and **edges**.

$$k_i = 6$$

$$k_j = 4$$



The number of connections a node has is its **degree** k .

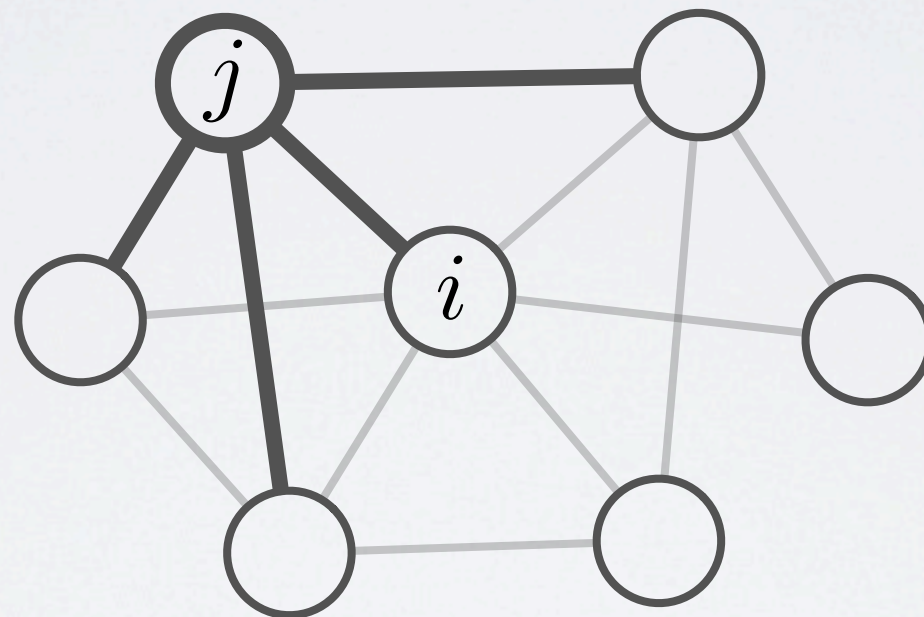
NETWORKS

Network analysis is a highly interdisciplinary research field in its own right.

Networks consist of **nodes** and **edges**.

$$k_i = 6$$

$$k_j = 4$$



The number of connections a node has is its **degree** k .
This abstract framework allows us to examine a wide range of networks with the same tools.

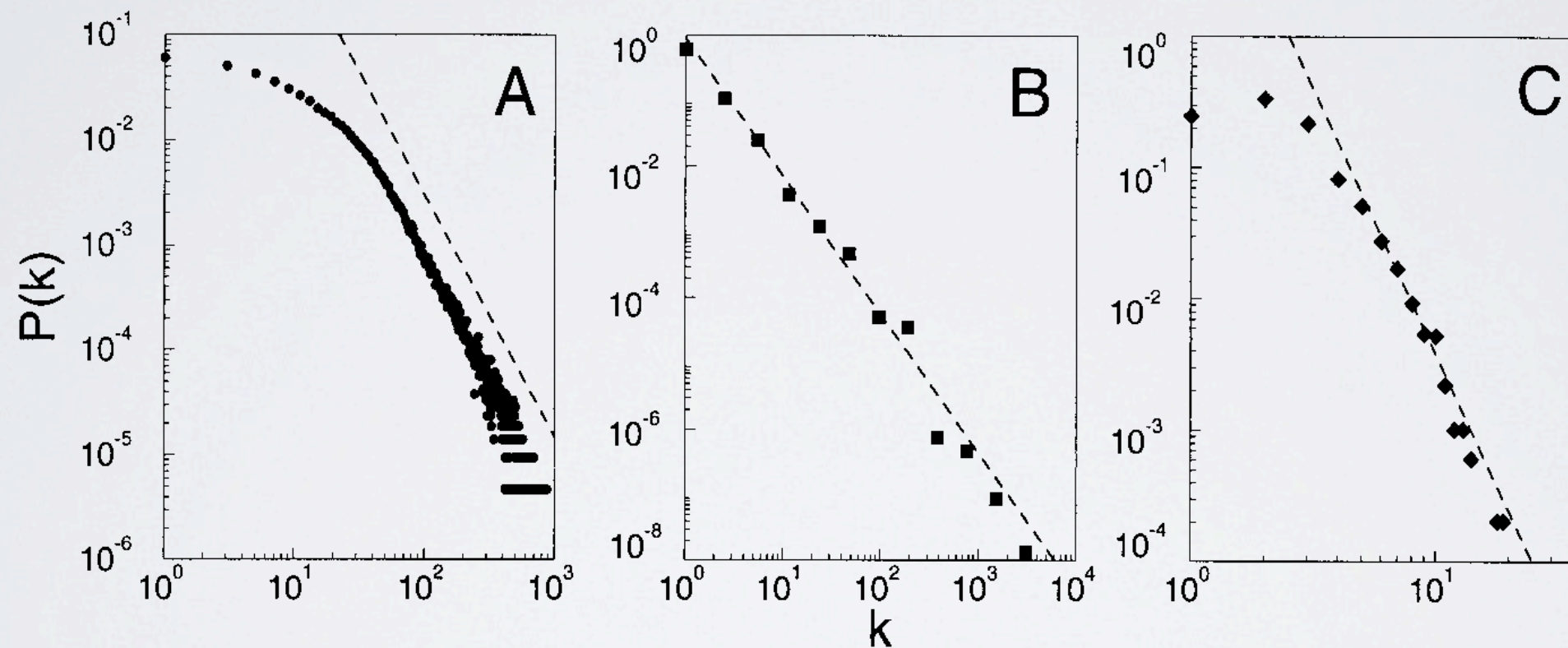
NETWORK ANALYSIS

NETWORK ANALYSIS

Many real-world networks have similar properties, such as a scale-free degree distribution.

NETWORK ANALYSIS

Many real-world networks have similar properties, such as a scale-free degree distribution.

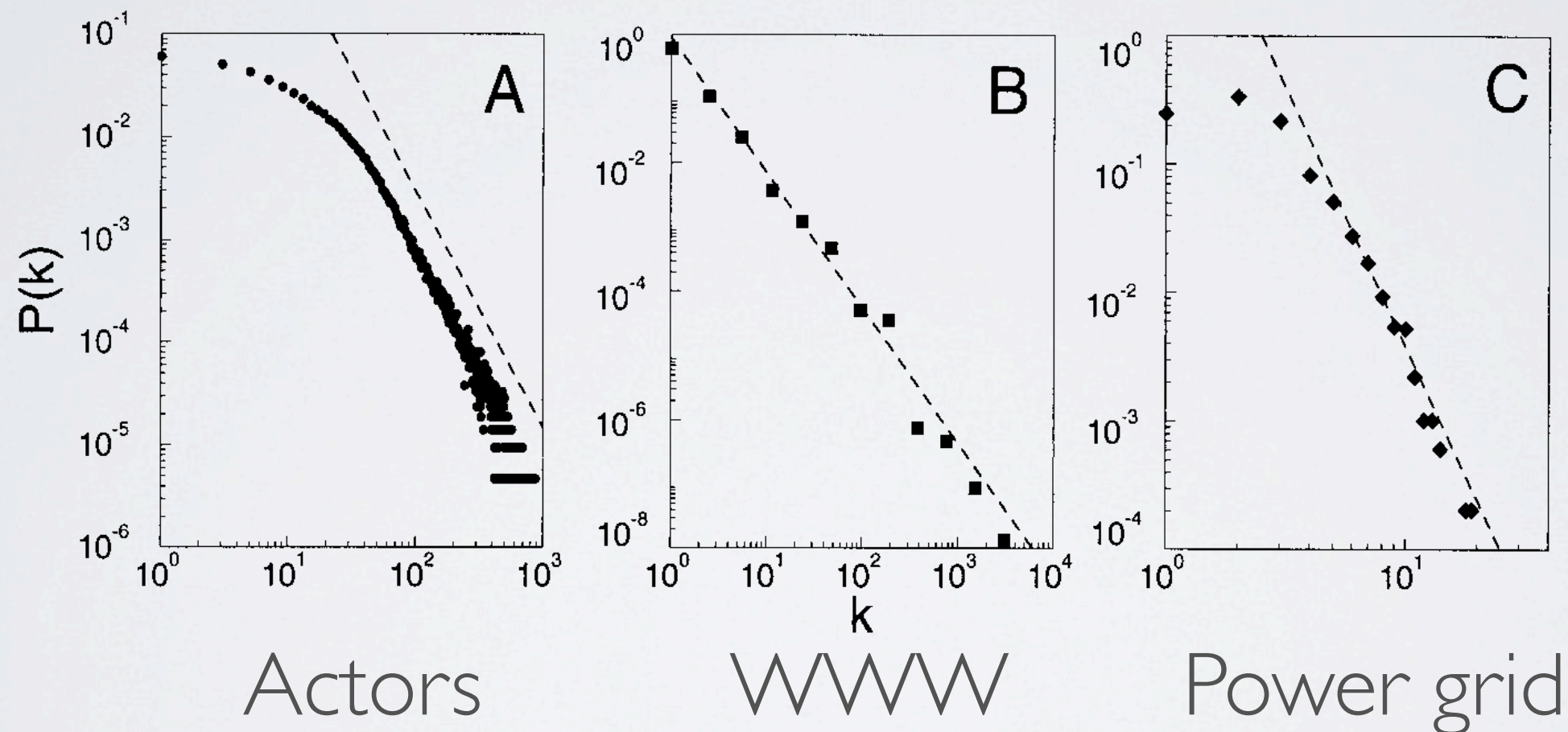


$$P(k) = k^{-\gamma}$$

Barabasi, A. & Albert, R. Emergence of scaling in random networks. Science 286, 509–512 (1999).

NETWORK ANALYSIS

Many real-world networks have similar properties, such as a scale-free degree distribution.



WHAT CAN WE MEASURE?

WHAT CAN WE MEASURE?

We can measure the characteristics of nodes and edges in many different ways.

WHAT CAN WE MEASURE?

We can measure the characteristics of nodes and edges in many different ways.

An obvious one is the **degree**. High-degree nodes, also called hubs, are individuals that have many social connections.

WHAT CAN WE MEASURE?

We can measure the characteristics of nodes and edges in many different ways.

An obvious one is the **degree**. High-degree nodes, also called hubs, are individuals that have many social connections.

A more subtle one is **betweenness**. For any pair of nodes in the network we can find the shortest path between them. The betweenness of a node measures how many shortest paths in the network pass through that node.

WHAT CAN WE MEASURE?

We can measure the characteristics of nodes and edges in many different ways.

An obvious one is the **degree**. High-degree nodes, also called hubs, are individuals that have many social connections.

A more subtle one is **betweenness**. For any pair of nodes in the network we can find the shortest path between them. The betweenness of a node measures how many shortest paths in the network pass through that node.

We can also study the **robustness** of the network by examining how the properties of a network change if nodes or edges are removed.

OUR DATA SETS

Our work focuses on two datasets:

OUR DATA SETS

Our work focuses on two datasets:

1) A network of letters written by Protestants during the Catholic reign of Mary I (1553-58).

OUR DATA SETS

Our work focuses on two datasets:

1) A network of letters written by Protestants during the Catholic reign of Mary I (1553-58).

2) Correspondence in



EXTRACTING THE NETWORK

EXTRACTING THE NETWORK

EXTRACTING THE NETWORK

Starting from manuscripts we can extract what might be called the *metadata* of the letters:

EXTRACTING THE NETWORK

Starting from manuscripts we can extract what might be called the *metadata* of the letters:

- sender
- recipient
- time
- place

EXTRACTING THE NETWORK

Starting from manuscripts we can extract what might be called the *metadata* of the letters:

- sender
- recipient
- time
- place

But from the letter content we can also extract:

- commendations
- mentioned links
- familial relationships
- letter carriers

PROTESTANT NETWORK

PROTESTANT NETWORK

We applied this approach to a network of letters written by Protestants during the Catholic reign of Mary I (1553-58).

PROTESTANT NETWORK

We applied this approach to a network of letters written by Protestants during the Catholic reign of Mary I (1553-58).

These letters reflect the persecution and martyrdom of Protestants during this period.

PROTESTANT NETWORK

We applied this approach to a network of letters written by Protestants during the Catholic reign of Mary I (1553-58).

These letters reflect the persecution and martyrdom of Protestants during this period.

From 289 letters we put together a network of 377 individuals and 795 edges. An edge denotes a social interaction, such as a letter being written, or a commendation.

PROTESTANT NETWORK

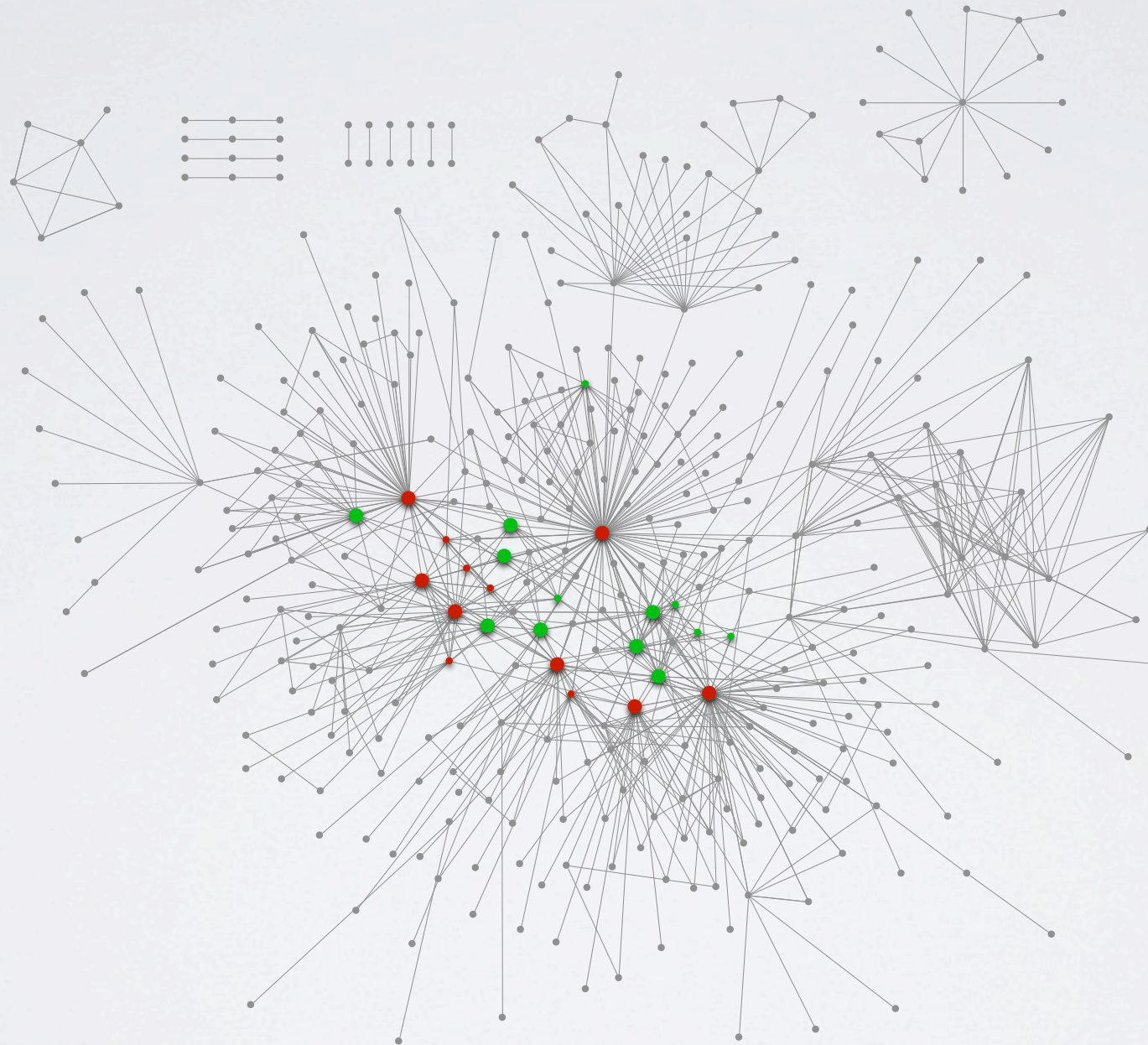
We applied this approach to a network of letters written by Protestants during the Catholic reign of Mary I (1553-58).

These letters reflect the persecution and martyrdom of Protestants during this period.

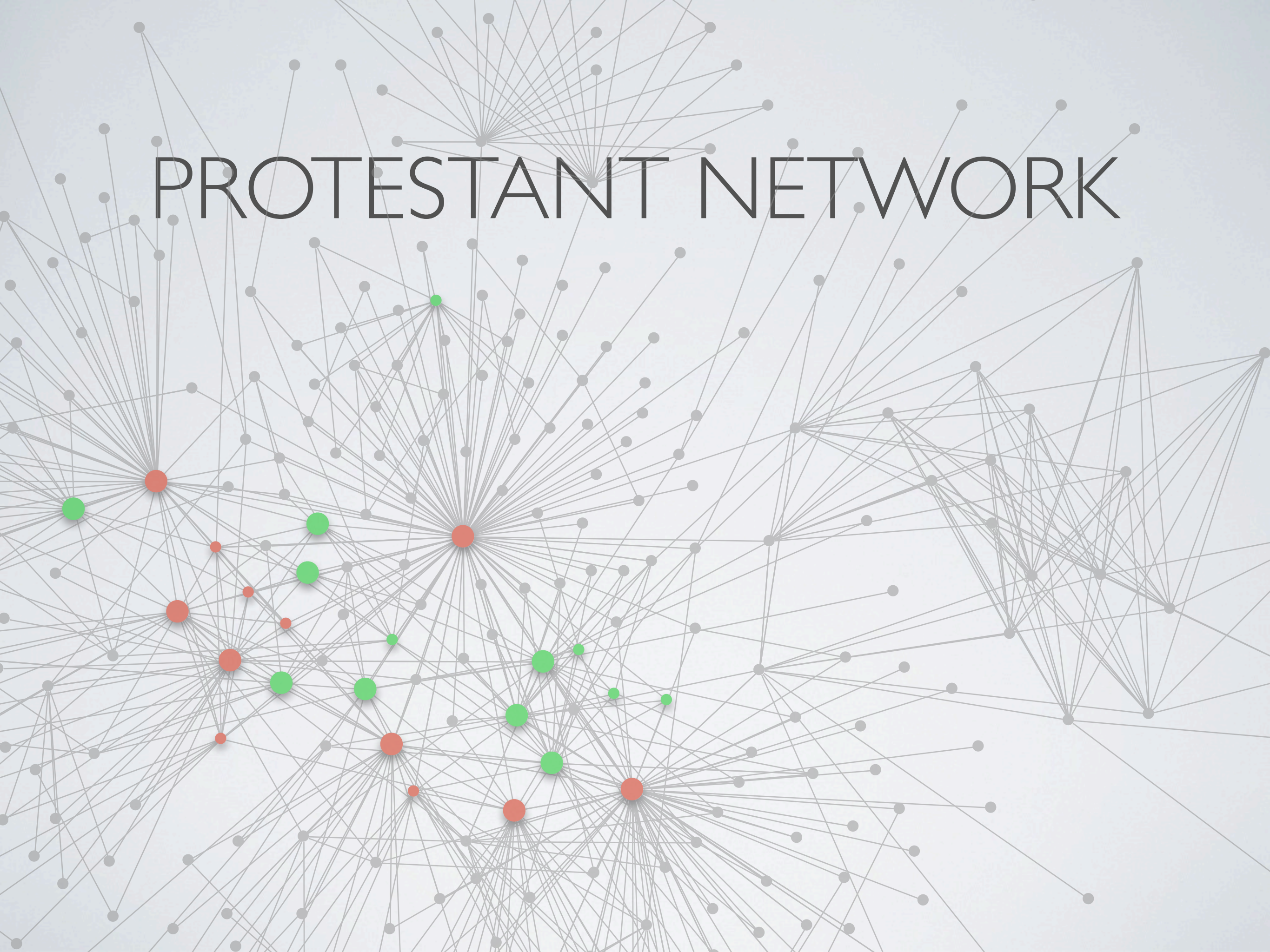
From 289 letters we put together a network of 377 individuals and 795 edges. An edge denotes a social interaction, such as a letter being written, or a commendation.

A subset of these edges is the actual correspondence network, which can be analysed separately.

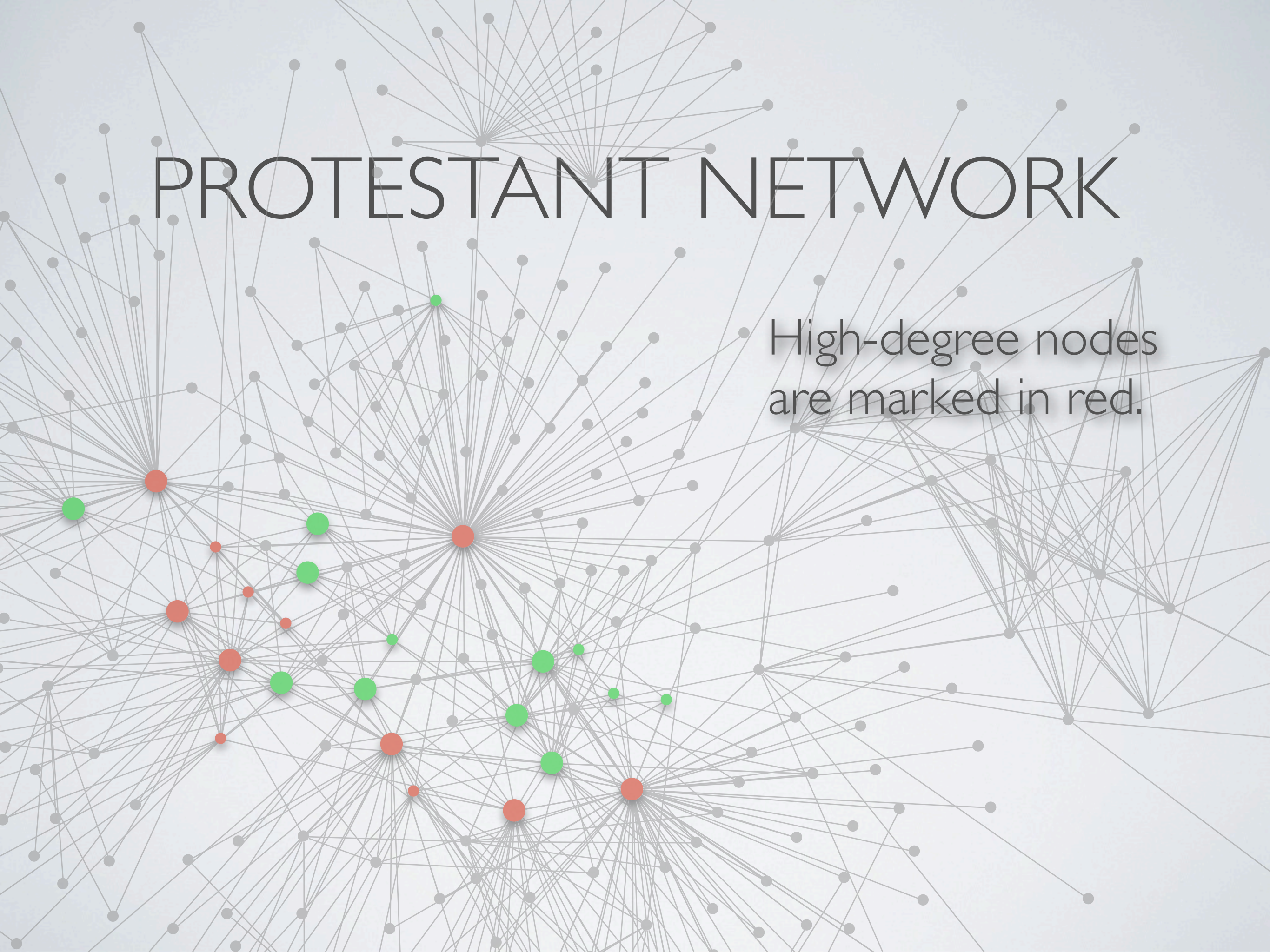
PROTESTANT NETWORK



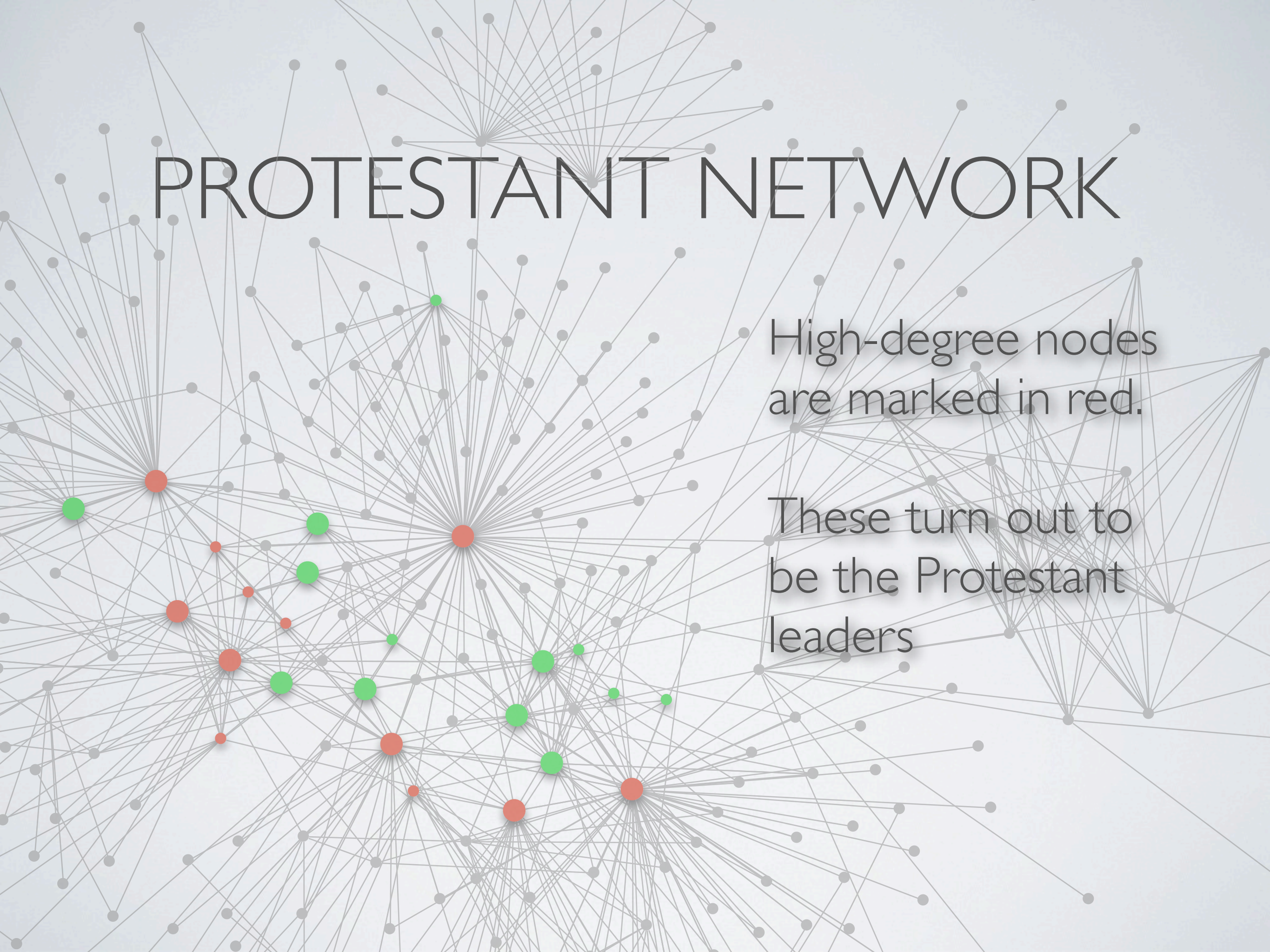
PROTESTANT NETWORK



PROTESTANT NETWORK

The image displays a large, intricate network graph. It consists of numerous nodes, represented by small circles, connected by a dense web of thin, grey lines representing edges. The nodes are distributed across the frame, with some areas showing higher concentrations and more connections than others. Several nodes are highlighted in red, and a few are highlighted in green, as indicated by the text on the right. The overall background is a light grey, and the network structure is the primary focus.

High-degree nodes
are marked in red.

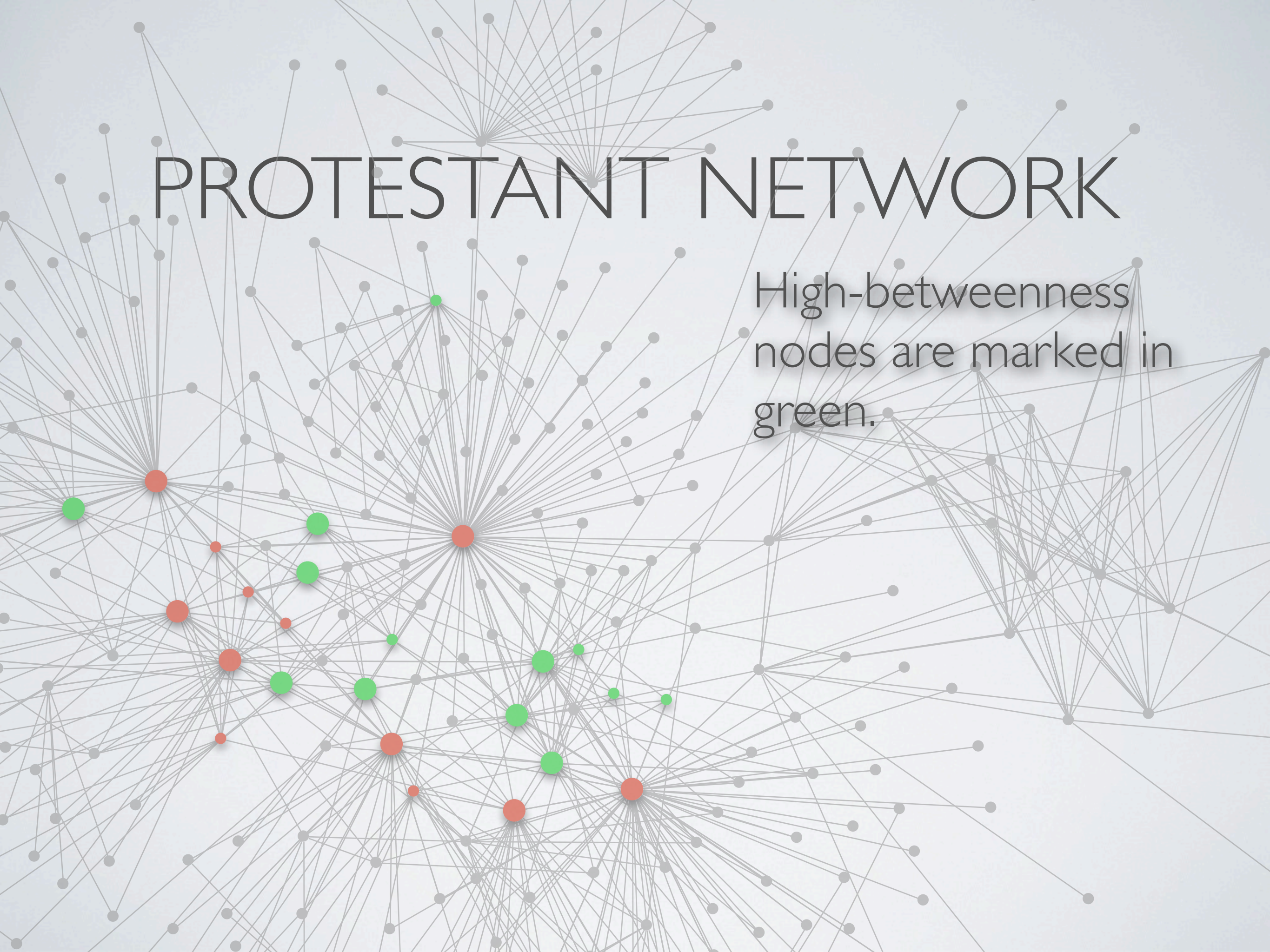


PROTESTANT NETWORK

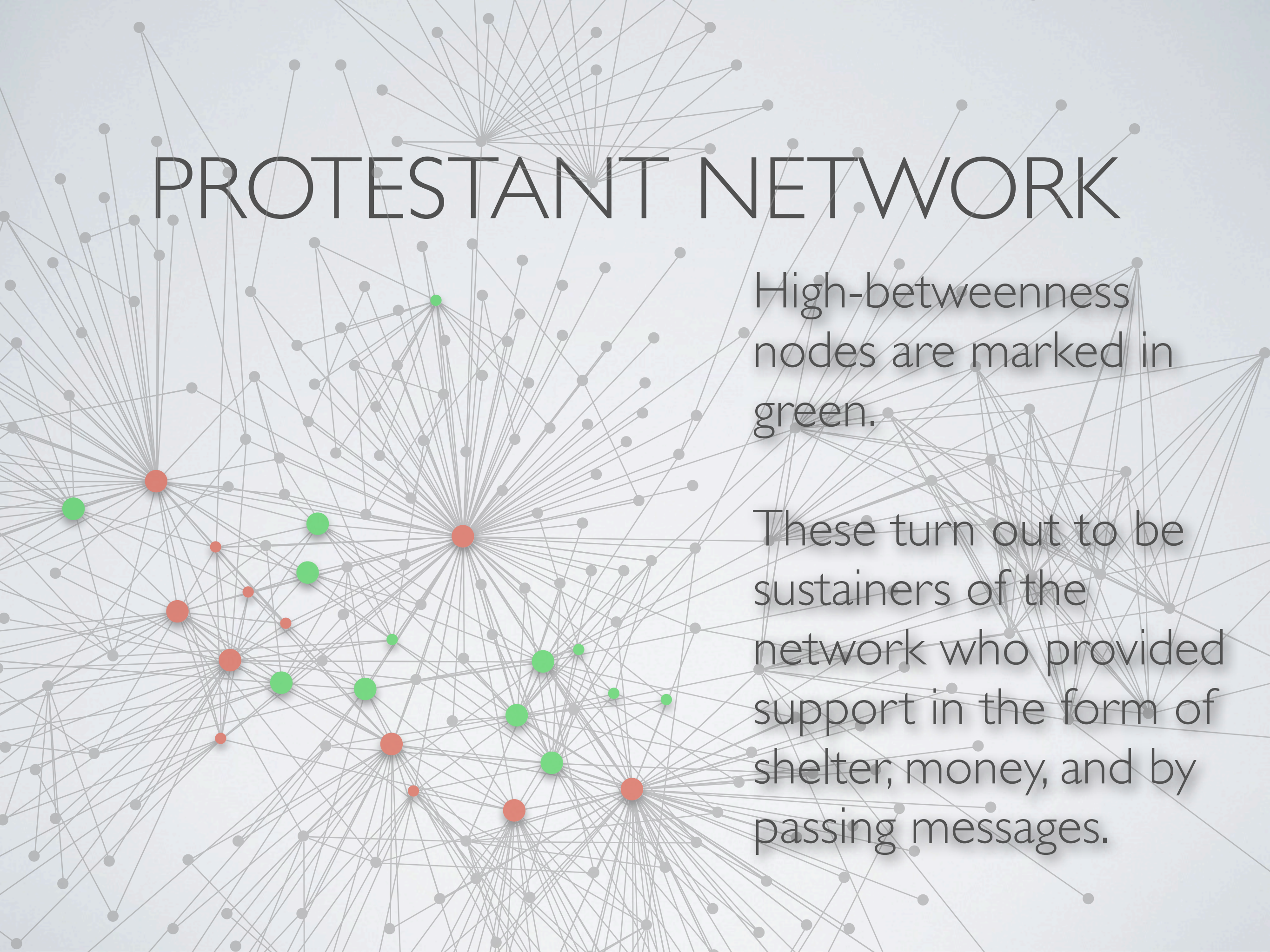
High-degree nodes
are marked in red.

These turn out to
be the Protestant
leaders

PROTESTANT NETWORK

A complex network graph visualization with many nodes and edges. The nodes are represented by small circles, and the edges are thin lines connecting them. The network is dense and spread across the frame. Some nodes are highlighted in green and red, indicating specific characteristics or roles within the network.

High-betweenness
nodes are marked in
green.



PROTESTANT NETWORK

High-betweenness nodes are marked in green.

These turn out to be sustainers of the network who provided support in the form of shelter, money, and by passing messages.

RESEARCH QUESTION

RESEARCH QUESTION

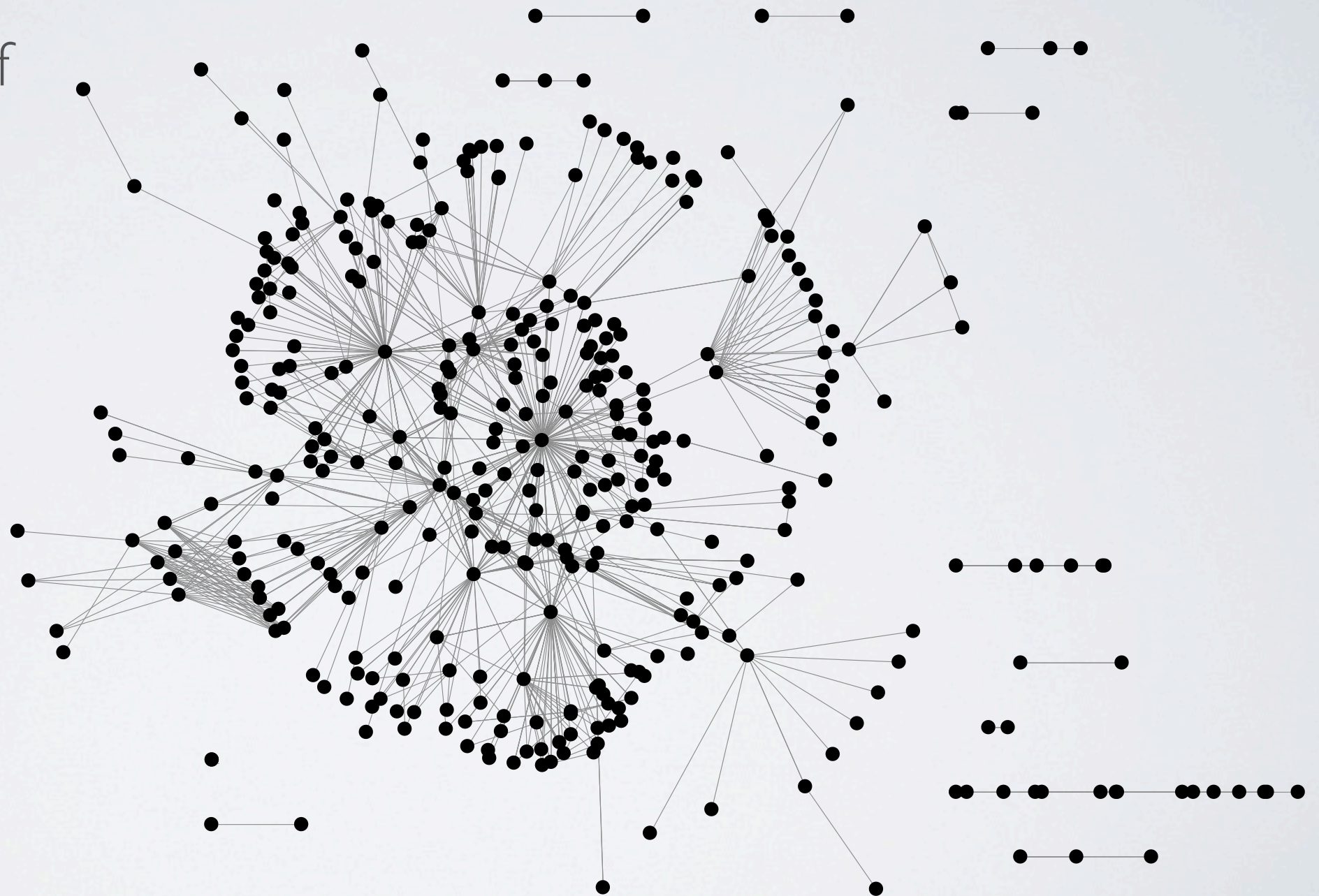
Research question for WG2:

How can letter network data identify social roles of individuals in a network, as leaders, sustainers, carriers, rebels, dissenters, informants, etc..?

ROBUSTNESS

ROBUSTNESS

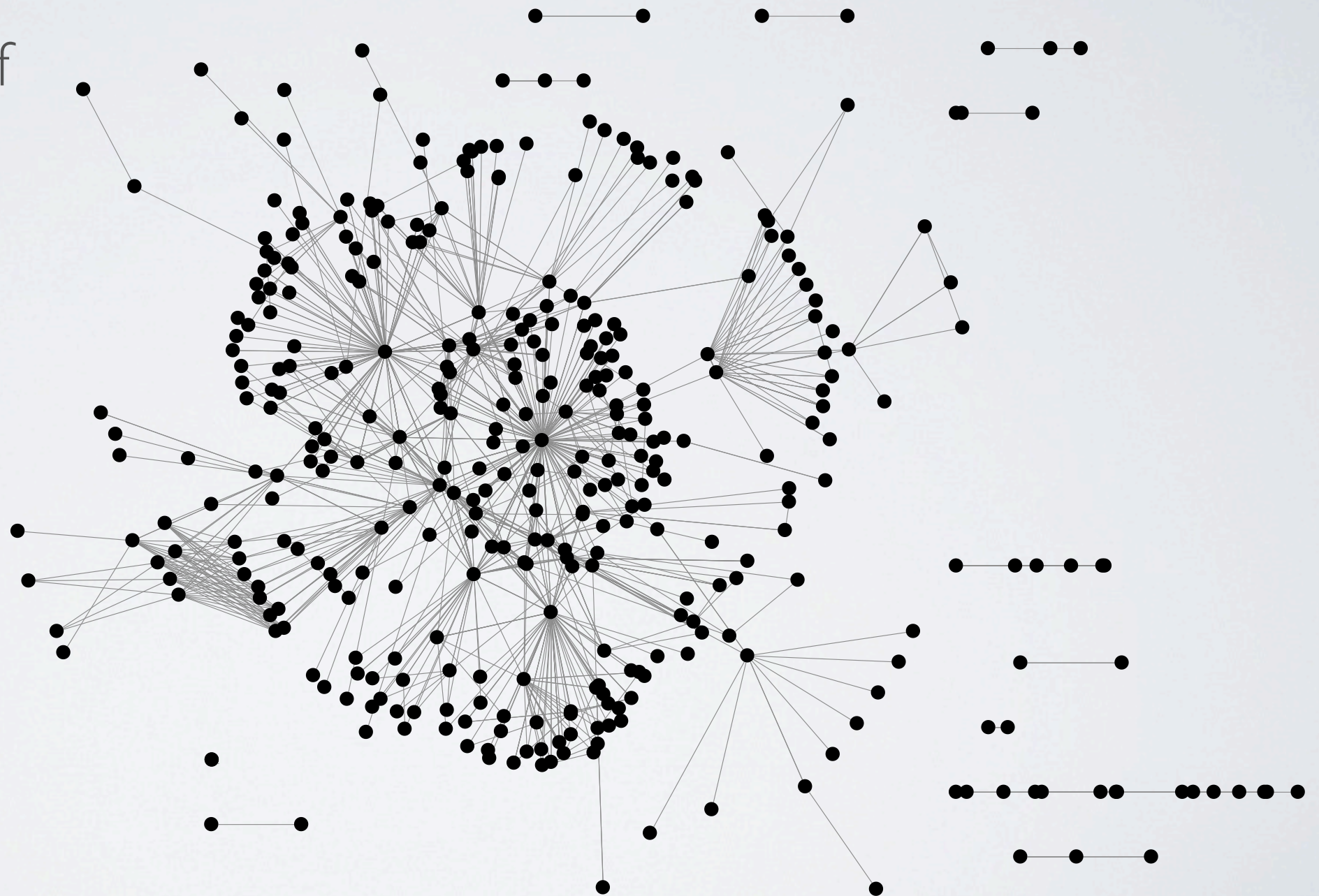
During the reign of Mary I many Protestants were executed.



ROBUSTNESS

During the reign of Mary I many Protestants were executed.

This campaign focused on the leaders of the movement, which were the hubs in the network.



ROBUSTNESS

During the reign of Mary I many Protestants were executed.

This campaign focused on the leaders of the movement, which were the hubs in the network.



ROBUSTNESS



ROBUSTNESS

But despite a large number of executions, the network does not fragment.



ROBUSTNESS

But despite a large number of executions, the network does not fragment.

This is because the sustainers keep it connected.



ROBUSTNESS



ROBUSTNESS

One of the best ways of fragmenting a network is in fact not to remove hubs, but to remove nodes or edges with high betweenness.



RESEARCH QUESTION

RESEARCH QUESTION

Research question for WG2:

How can letter network data examine temporal changes in a network, both on a local and global scale, particularly in the face of political and religious change?

STATE PAPERS ONLINE

STATE PAPERS ONLINE

Our next project is a much larger data set:

- The State Papers for the entire 16th century in England.
- These are the official government records.
- 130,000 items of correspondence connecting ~30,000 individuals.
- Many of these items connect people across Europe too.
- This data covers the reigns of four monarchs, both Catholic and Protestant.

DISAMBIGUATION

DISAMBIGUATION

To create a network of individuals we need to first make sure whether:

- a given named individual is referred to by any other name or title in another document
- a given name or title can refer to another individual in another document

Examples:

- *de la Motte* and *de la Mote* might be the same individual or not.
- *Archbishop of Canterbury* can refer to several different people, depending on the date of the letter.

DISAMBIGUATION

To create a network of individuals we need to first make sure whether:

- a given named individual is referred to by any other name or title in another document
- a given name or title can refer to another individual in another document

Examples:

- *de la Motte* and *de la Mote* might be the same individual or not.
- *Archbishop of Canterbury* can refer to several different people, depending on the date of the letter.

80% of Data Science is Data Cleaning!

DISAMBIGUATION

DISAMBIGUATION

DISAMBIGUATION

The disambiguation process also offers opportunities:

DISAMBIGUATION

The disambiguation process also offers opportunities:

The creation of unique identifiers is a prerequisite for linked data. Efforts like this one can serve as nuclei for linked data projects.

DISAMBIGUATION

The disambiguation process also offers opportunities:

The creation of unique identifiers is a prerequisite for linked data. Efforts like this one can serve as nuclei for linked data projects.

We can use the disambiguation name lists and mappings to search letter content and extract mentions of individuals and further social links.

RESEARCH QUESTION

RESEARCH QUESTION

Research question for WG2:

How in general can disambiguation of letter networks drive linked data efforts and content analysis?

BIAS

BIAS

Letter collections come with inevitable biases that reflect their historical origins. This should be viewed as an opportunity rather than a problem, as the distortion of quantitative results by these biases can bring historical facts to the surface.

BIAS

Letter collections come with inevitable biases that reflect their historical origins. This should be viewed as an opportunity rather than a problem, as the distortion of quantitative results by these biases can bring historical facts to the surface.

Edward Courtenay is unusually well-connected in the State Papers Domestic letter network for Mary I reign, and is therefore highlighted in quantitative network analysis results. This is not a direct reflection of his importance, but an indirect one. Courtenay was in exile, but his correspondence was intercepted as he was deemed potentially dangerous.

RESEARCH QUESTION

RESEARCH QUESTION

Research question for WG2:

How can biases in the data be linked to the history of both individuals and collections?

CONCLUSION

CONCLUSION

These are some of the research questions that WG2 could look at.

CONCLUSION

These are some of the research questions that WG2 could look at.

Some of these are straightforward questions linking established historical questions about individuals and communities to quantitative measurements. This is an iterative process requiring a close dialogue between traditional scholarship and a quantitative bird's-eye view.

CONCLUSION

These are some of the research questions that WG2 could look at.

Some of these are straightforward questions linking established historical questions about individuals and communities to quantitative measurements. This is an iterative process requiring a close dialogue between traditional scholarship and a quantitative bird's-eye view.

Other research questions relate to the challenges of network analysis in the humanities, namely that any data set will be flawed in terms of both accuracy and bias.

CONCLUSION

These are some of the research questions that WG2 could look at.

Some of these are straightforward questions linking established historical questions about individuals and communities to quantitative measurements. This is an iterative process requiring a close dialogue between traditional scholarship and a quantitative bird's-eye view.

Other research questions relate to the challenges of network analysis in the humanities, namely that any data set will be flawed in terms of both accuracy and bias.

It is important to realise that the process of cleaning the data can be a valuable step towards integration with other databases, and that biases are part of the history.