Modelling Influence Spreading on Complex Networks

Vesa Kuikka
Modelling Influence Spreading on Complex Networks

Vesa Kuikka

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall 240 L2 in Puunjalostustekniikka 1 on 22th November 2022 at 12 noon.

Aalto University
School of Science
Department of Computer Science
Complex Systems
Supervising professor
Professor Jouko Lampinen, Aalto University

Thesis advisor
Professor Kimmo K. Kaski, Aalto University

Preliminary examiners
Associate professor, Takashi Shimada, The University of Tokyo, Japan
Associate professor, János Török, Budapest University of Technology and Economics, Hungary

Opponent
Professor Martin Rosvall, Umeå University, Sweden
Abstract

Network Science is a growing field of research and it has gained research interest in many application areas. One definition of network science is the study of network representations of physical, biological and social phenomena aiming at predictive models of these phenomena. In addition to specialised methods of network modelling, networks can be studied with general methods of computer science, statistics and many other fields of mathematics.

In this thesis I have developed an influence spreading model that can be applied to social networks, epidemic spreading, and spreading processes in physical and biomedical networks. The model is based on detailed network structures of nodes, links and paths in the network. Individual node and link weights are interpreted as probabilities of transferring information, influence or infection over nodes and via links, respectively.

The model is demonstrated with social networks, epidemic spreading in structured organisations and interoperability in brain networks. In this context, I have proposed a community detection method to discover groups or connected regions in network structures, especially in social networks. In the case of brain networks, I have also discussed how the influence spreading model can be exchanged with a network connectivity model. This demonstrates how the community detection method can be used with different network models. The proposed community detection method and novel metrics are based on the influence spreading model presented in this study.

The community detection method detects overlapping and hierarchical structures. Node and link weights can be used to describe weak and strong interactions or different levels of granularity of results. The method searches the local maxima of an objective function in detecting different splittings of the network. The objective function that I have used in the algorithm for detecting communities and sub-communities can be used as a quality function to compare the cohesion between any sets of nodes in the network structure. In addition, I have proposed alternative quality functions for measuring the probability of formation and robustness of composition for different community structures. Different probabilistic temporal distributions can be implemented in the spreading model. Temporal spreading has been demonstrated with two different distributions, the Poisson distribution and a probabilistic distribution describing temporal delays in message forwarding events.

In order to investigate epidemic spreading in structured organisations using the influence spreading model I have used two main versions: complex contagion and simple contagion. The complex contagion model allows loops and breakthrough infection via nodes. The simple contagion model does not allow loops or breakthrough infection. It is assumed that the complex contagion model can describe the spread of virus infections where the main method of spreading is through the air with droplets and particles and breakthrough infection has a large effect.

Keywords Influence spreading model, Social network, Complex network
Tekijä
Vesa Kuikka

Väitöskirjan nimi
Modelling Influence Spreading on Complex Networks

Julkaisija
Perustieteiden korkeakoulu

Yksikkö
Tietotekniikan osasto

Sarja
Aalto University publication series DOCTORAL THESES 142/2022

Tutkimusala
Kompleksiset systemit

Käsikirjoituksen pvm
20.04.2022

Väitöspäivä
22.11.2022

Väittelyluvan myöntämispäivä
05.10.2022

Kieli
Englanti

Monografia

Artikkeliväitöskirja

Esseeväitöskirja

Tiivistelmä

Verkostotiede on kasvava tutkimusalue ja se on saavuttanut mielenkiintoa monilla sovellusalueilla. Verkostotiede voidaan määritellä tieteen alana, joka tutkii fysiikaalisia, biologisia ja sosiaalisia verkostoja ja pyrkii kuvaamaan ja ennustamaan erilaisia ilmiöitä verkostomallien avulla. Verkostojen on mahdollista tutkia myös muilla yleisillä tietojenkäsittelytieteen ja tilastotieteen menetelmillä.


Avainsanat
Vaikutuksen leviäismalli, Sosiaalinen verkosto, Kompleksinen verkosto

ISBN (painettu) 978-952-64-0965-8
ISBN (pdf) 978-952-64-0966-5

ISSN (painettu) 1799-4934
ISSN (pdf) 1799-4942

Julkaisupaikka Helsinky
Painopaikka Helsinky
Vuosi 2022

Sivumäärä 261

Acknowledgements

I would first like to thank my supervisor professor Jouko Lampinen from Aalto University and my instructor professor emeritus Kimmo K. Kaski from Aalto University for the guidance during my doctoral studies.

I would like to acknowledge my co-authors Daniel Monsivais and Minh An Antti Pham for their contributions to the research articles used as sources in this study.

I am grateful to the preliminary examiners Dr Takashi Shimada from the University of Tokyo and Dr János Török from the Budapest University of Technology and Economics for their valuable comments on an earlier version of the manuscript.

Helsinki, Finland, August 25th, 2022,

Vesa Kuikka
# Contents

Acknowledgements

List of Symbols

List of Abbreviations

List of Publications

Author’s Contribution

Other Publications

List of Figures

List of Tables

1. Introduction ........................................................................ 1
   1.1 Network Science ......................................................... 2
   1.2 Modelling Social Networks .......................................... 4
   1.3 Basic Concepts of Network Modelling ........................................ 4
   1.3.1 Structural Holes ............................................................. 5
   1.3.2 Weak Ties ........................................................................... 6
   1.3.3 Complex Contagion ...................................................... 6
   1.3.4 Processes on complex networks ...................................... 7

2. Research Objectives ..................................................... 9
   2.1 Research Question ....................................................... 9
   2.2 Delimitations ........................................................................ 10
   2.3 Research Gaps ........................................................................... 10

3. Related Work ...................................................................... 13
   3.1 Social Influence Analysis .............................................. 13
   3.2 Centrality and Betweenness Measures ........................... 14
   3.3 Network Models and Community Formation .................. 16
   3.4 Community Detection ...................................................... 17
   3.5 Epidemic Spreading Models .......................................... 18
   3.5.1 SIS and SIR models ....................................................... 19
   3.5.2 Selected Epidemic Spreading Models Related to this Work 20

4. Influence Spreading Model ............................................. 21
   4.1 Influence Spreading Model ........................................... 21
4.1.1 Properties of the Model ......................................................... 21
4.1.2 Definition of the Model ........................................................ 22
4.2 Scalability of the Influence Spreading Model ......................... 24
4.3 Centrality and Betweenness Measures ..................................... 25
4.3.1 Definitions Based on the Influence Spreading Matrix .......... 25
4.3.2 Comparison with Other Centrality Measures ....................... 26
4.4 Comparison of the Centrality and Betweenness Measures in a Small
  Ego-centric Network ........................................................................ 27
4.5 Community Detection .............................................................. 31
5. Applications of the Influence Spreading Model ...................... 35
  5.1 Social Networks – a Case Study .............................................. 35
  5.2 Social Media Network – an Example ..................................... 39
  5.3 Ego-centric Networks ............................................................. 41
  5.4 Brain Network ........................................................................ 43
  5.5 Epidemic Spreading ............................................................... 46
6. Scientific Results ...................................................................... 57
7. Future Research ....................................................................... 61
8. Conclusion .............................................................................. 63
Bibliography .................................................................................. 65

Publications .................................................................................... 69
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>Redundancy of an ego-centric network in the Burt’s model in Eq. 1</td>
</tr>
<tr>
<td>$t$</td>
<td>Number of connections in a network in the Burt’s model in Section 1.2.1</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of neighbouring nodes excluding ego in the Burt’s model</td>
</tr>
<tr>
<td>$e_G$</td>
<td>Effectiveness of node $G$ in the Burt’s model in Eq. 2</td>
</tr>
<tr>
<td>$C_i^{(clos1)}$</td>
<td>Standard definition of closeness centrality measure in Eq. 3</td>
</tr>
<tr>
<td>$C_i^{(clos2)}$</td>
<td>Alternative definition for the standard closeness centrality measure in Eq. 4</td>
</tr>
<tr>
<td>$C_i^{(betw)}$</td>
<td>Standard definition of betweenness measure in Eq. 5</td>
</tr>
<tr>
<td>$d_{i,j}$</td>
<td>Distance (or the number of links) between nodes $i$ and $j$</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of nodes in a network</td>
</tr>
<tr>
<td>$\sigma_{s,t}$</td>
<td>Number of shortest paths from source node $s$ to target node $t$</td>
</tr>
<tr>
<td>$\sigma_{s,t}(i)$</td>
<td>Number of shortest paths from source node $s$ to target node $t$ including node $i$</td>
</tr>
<tr>
<td>$C_i^{(Katz)}$</td>
<td>Katz centrality measure in Eq. 6</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Decay or model parameter in Katz, authorities, hubs and PageRank centralities in Section 3.2</td>
</tr>
<tr>
<td>$A = (A_{i,j})$</td>
<td>Adjacency matrix</td>
</tr>
<tr>
<td>$C_i^{(auth)}$</td>
<td>Authorities centrality in Eq. 7</td>
</tr>
<tr>
<td>$C_i^{(hub)}$</td>
<td>Hubs centrality in Eq. 8</td>
</tr>
<tr>
<td>$C_i^{(PR)}$</td>
<td>PageRank centrality in Eq. 9</td>
</tr>
<tr>
<td>$P_{i,j}$</td>
<td>Probability of random walk from node $i$ to node $j$ in Eq. 10</td>
</tr>
<tr>
<td>$r(V)$</td>
<td>Reliability of network $V$ in Eq. 11</td>
</tr>
<tr>
<td>$\mathcal{O}$</td>
<td>Set of all connected states of a network in Eq. 11</td>
</tr>
</tbody>
</table>
$S$ \hspace{1cm} Set of links in Eq. 11

$p_e$ \hspace{1cm} Probability of a functioning link $e$ in the definition of $r(V)$ in Eq. 11

$M$ \hspace{1cm} Modularity of a network in Eqs. 12–13

$m$ \hspace{1cm} $2m$ is the number of stubs in a network in Eq. 12

$k_v$ \hspace{1cm} Node degree of node $v$ in Eq. 12

$s_v$ \hspace{1cm} Variable indicating whether node $v$ is a member (or is not a member) of a community in Eq. 12

$B$ \hspace{1cm} Modularity matrix, $B_{v,w} = A_{v,w} - (2m)^{-1} k_v k_w$ in Eq. 13

$t$ \hspace{1cm} Time

$I(t)$ \hspace{1cm} Fraction of infected individuals at time $t$ in Eqs. 14–15

$R(t)$ \hspace{1cm} Fraction of recovered individuals at time $t$ in Eq. 15

$\beta$ \hspace{1cm} Infection rate in Eq. 14

$\mu$ \hspace{1cm} Recovery rate in Eqs. 14–15

$\lambda$ \hspace{1cm} Transition rate in Eq. 15

$\langle k \rangle$ \hspace{1cm} Expected value of node degree $k$

$C(s, t)$ \hspace{1cm} Element $s, t$ of the influence spreading matrix $C$

$L$ \hspace{1cm} Path length

$L_{\text{max}}$ \hspace{1cm} Maximum path length

$\nu$ \hspace{1cm} Number of visits on a node

$L$ \hspace{1cm} List of paths

$L(i)$ \hspace{1cm} Path $i$ in the list of paths

$L(L(i))$ \hspace{1cm} Path length of path $L(i)$

$S_{L(L(i))}(T)$ \hspace{1cm} Time-dependent factor in the spreading process at time $T$

$W_{L(i)}$ \hspace{1cm} Product of link and node weights along the path

$w_{st}$ \hspace{1cm} Link weight from node $s$ to node $t$

$w_n$ \hspace{1cm} Node weight of node $n$

$P_{L(L(i))}(T)$ \hspace{1cm} Probability of spreading at time $T$

$L(i_1) \cap L(i_2)$ \hspace{1cm} Common part of paths $L(i_1)$ and $L(i_2)$ before the first branching of the two paths

$N_L$ \hspace{1cm} Number of paths with path lengths $L \leq L_{\text{max}}$
\( S_L(T) \) Survival function value with path length \( L \) at time \( T \)

\( \lambda \) Intensity parameter of the Poisson distribution in Eq. 18

\( C_s^{(out)} \) Out-centrality of node \( s \) in Eq. 19

\( C_t^{(in)} \) In-centrality of node \( t \) in Eq. 20

\( b_n \) Betweenness centrality of node \( n \) in Eq. 21

\( C \) Coherence of a network in Eq. 22

\( B_n \) Coherence of a network with node \( n \) removed in Eq. 23

\( w_l \) Link weight of link \( l \)

\( P \) Objective function for community detection in Eq. 24

\( n_{ij} \) Number of communication events from node \( i \) to node \( j \) in Eq. 25

\( \alpha \) Tuning parameter of the link strength function in Eq. 25

\( q \) Model parameter for preventive measure ‘Leadership’

\( a \) Model parameter for preventive measure ‘Peers or roommates’

\( r \) Model parameter for preventive measure ‘Same department’

\( z \) Model parameter for preventive measure ‘Other contacts’

\( R_0 \) Basic reproduction number

\( r_0 \) Reproduction number in a structured organisation

\( W_l \) Model parameter describing link strengths with no preventive measures
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>Complex contagion</td>
</tr>
<tr>
<td>SC</td>
<td>Simple contagion</td>
</tr>
<tr>
<td>Out-C</td>
<td>Out-centrality</td>
</tr>
<tr>
<td>In-C</td>
<td>In-centrality</td>
</tr>
<tr>
<td>Alg-cc</td>
<td>Analytic algorithm for complex contagion spreading</td>
</tr>
<tr>
<td>Alg-sacc</td>
<td>Analytic algorithm for complex contagion spreading with self-avoiding paths</td>
</tr>
<tr>
<td>Sim-sc</td>
<td>Simulation algorithm for simple contagion spreading</td>
</tr>
</tbody>
</table>
List of Publications

This doctoral dissertation consists of a summary and the following publications which are referred to in the text by their numerals.


Author’s Contribution

P I. Influence spreading model used to analyse social networks and detect sub-communities.

The sole work of the author.

P II. Modelling community structure and temporal spreading on complex networks.

The sole work of the author.

P III. Terrorist network analyzed with an influence spreading model.

The sole work of the author.

P IV. Models of influence spreading on social networks.

VK was the main author of the paper. VK came up with the main idea and designed the experiments. The contribution of AP to the writing of the paper was about 20 per cent. AP implemented the simulation software (about 60 per cent of the program code).

P V. Influence spreading model in analysing ego-centric social networks.

VK, DM and KK contributed equally to the writing of the paper. DM was responsible for the statistical analysis and visualization of the data.

P VI. Subsystem cooperation in complex networks – Case brain network.

The sole work of the author.

P VII. Modelling epidemic spreading in structured organisations.

The sole work of the author.
Other Publications

These publications are supplementary articles.


List of Figures

**Figure 1.** Ego-centric network structure adapted from [1].

**Figure 2.** Five alternative centrality measure values for the seven nodes \( \{A, B, C, D, E, F \text{ and } G\} \) of the ego-centric network in Fig. 1. Results are shown for link weight values \( w_l = 0.05 \) (top) and \( w_l = 0.5 \) (bottom). Also, the average values of the five measures are shown. Out-C, CC and In-C, CC are the out-centrality and in-centrality values when loops and breakthrough effects are allowed during the spreading of influence on the network. C-SC show the in- and out-centrality values for self-avoiding paths and \( L = 1 \) show the centrality values with the limited path length. Notice that \( L = 1 \) values are proportional to the degree values of nodes. For comparison, I show also the effective centrality values of Burt’s model [1].

**Figure 3.** Betweenness values for the seven nodes \( \{A, B, C, D, E, F \text{ and } G\} \) of the ego-centric network of Fig. 1 as a function of the link weight value \( w_l \). The top figure shows the results when loops are allowed and the bottom figure shows the results for self-avoiding paths.

**Figure 4.** This simple network structure is used to illustrate three alternative spreading models in Table 4.

**Figure 5.** The results calculated with the community analysis method of this study show 11 strong divisions of the Les Misérables network. The borders of the divisions show building blocks of the community structure (P II).

**Figure 6.** Structure of the terrorist network [44, 45] with detected sub-communities in the influence spreading model.

**Figure 7.** Structure of the terrorist network [44, 45] with detected sub-communities in the influence spreading model.

**Figure 8.** The values of closeness centrality measure of the nodes in the terrorist network in Fig. 7.

**Figure 9.** The values of betweenness centrality measure of the nodes in the terrorist network in Fig. 7.
Figure 10. The ratios of the results with the weighted and equal weights are calculated from Figs. 8—9. Closeness centrality is shown by bars (left vertical axis) and betweenness centrality with dots (right vertical axis).

Figure 11. The values of the link betweenness centrality for the 50 most important links in the terrorist network of Fig. 7.

Figure 12. A Facebook social media network of 4039 nodes. Nodes represent users in the network and links represent connections between them. Node sizes in the graph describe the centrality values of nodes.

Figure 13. The values of out-centrality and in-centrality measures of the Facebook social media network of Fig. 12 with link weights $w_l = 0.05$ for Alg-cc and Sim-sc. Alg-cc results are displayed for the maximum path lengths $L_{\text{max}} = 8$, $L_{\text{max}} = 11$ and $L_{\text{max}} = 13$. Alg-cc show complex contagion results and Sim-sc show simple contagion results. The order of source nodes (horizontal axis) is determined in the ascending order of Sim-sc out-centrality or in-centrality values.

Figure 14. The normalised influence in-centrality (on the left) and out-centrality (on the right) values for males (blue) and females (red) as a function of the ego’s age and the link weight model in Eq. 25 with parameter $\alpha = 0.01$ (top) and $\alpha = 0.5$ (bottom). The curves for $\alpha = 0.01$ and $\alpha = 0.5$ represent interaction and connectivity facets respectively.

Figure 15. Egos’ betweenness centrality values (Eq. 21) as a function of egos’ age for males (blue) and females (red). The averages are calculated for two different values of the link weight parameter $\alpha = 0.01$ and $\alpha = 0.5$ shown in the left and right panels. The curves for $\alpha = 0.01$ and $\alpha = 0.5$ represent interaction and connectivity facets respectively.

Figure 16. The values of the in-centrality and out-centrality measures for the 122 first regions (nodes) of the brain network.

Figure 17. The values of the betweenness measures for the 122 first regions (nodes) of the brain network.

Figure 18. The main division into two hemispheres is seen in the upper and lower parts of the figure. The second-order division is indicated by colours. Regions in detected subsystems that cooperate across the main division are indicated by larger font numbers (the figure is created by the Force Atlas layout of Gephi software).

Figure 19. The network view of connections in the organisation structure. Peer (or roommate) relations are shown with the black colour and other close relations with the blue colour. Nodes 13—14, 54—55, 95—96 and 136—137 are bridge
nodes between administrative staff and managers inside the four departments. Dashed lines show the borders of the administrative staff and the four departments.

**Figure 20.** Reproduction number $r_0$ as a function of the parameter value $W_l$ in the organisation of Fig. 19 for no preventive measures (—) and five different preventive measures.

**Figure 21.** Percentage of infected nodes as a function of the number of initial spreaders (#) in the organisation of Fig. 19 for no preventive measure (—) and contacts limited between departments (z). The parameter values of $W_l$ for the seven curves from top to bottom are $W_l = 0.22, 0.18, 0.14, 0.12, 0.10, 0.08$ and 0.06.

**Figure 22.** Percentage of infected nodes in the organisation of Fig. 19 for different path lengths (top) and the same path lengths with the preventive measure of limiting cross-department contacts (bottom).

**Figure 23.** Values of out-centrality, in-centrality and betweenness measures are shown as a function of the node numbers in Fig. 19. The parameter value of $W_l = 0.15$ is used in the calculations.

**Figure 24.** Community immunity without preventive measures (—) is achieved when about 48.1 % of the members of the organisation in Fig. 19 are immune and about 43 % with the preventive measure of limiting cross-department contacts (’z’). Out-centrality (dashed curves) and betweenness (dotted curves) show lower and upper bounds when these metrics are used for prioritising nodes in the network in Fig. 19.

**Figure 25.** Percentage of infected individuals for one initial spreader (index case) in the organisation of Fig. 19 for no preventive measures (—) and preventive measures ’z’, ’r’, ’a’, ’z&r’ and ’z&r&a’ in the complex contagion (CC) model (top) and the simple contagion (SC) model (bottom).
List of Tables

**Table 1.** Effective sizes and degrees of the nodes in Fig. 1.

**Table 2.** Running times for various real-world networks (A I) on a workstation with Intel Xeon X5690 processors.

**Table 3.** This table is a key to the structure of Table 4 and the content of the six matrices describing three algorithms and two values of the link weights.

**Table 4.** Numerical values from the three spreading models of the Alg-cc, Alg-sacc and Sim-sc algorithms on panes from top to bottom respectively. Link weight values $w_l = 0.05$ (left) and $w_l = 0.5$ (right) are used in the computations.

**Table 5.** Research questions **RQ1—RQ5** have been answered in Publications **P I—P VII.** In addition, I have indicated the most important formulas in the publications related to the research questions.
1. Introduction

In this thesis, I focus on investigating influence spreading and present a model that can be used in and extended to various applications to describe different spreading processes on network structures. Possible application areas are describing influence spreading in social networks, information transfer in communication networks, transportation in infrastructure networks and epidemic spreading in populations. Different spreading processes can exist in these applications which should be considered in the spreading model. Detailed level descriptions of the network and its elements are needed in calculating various quality measures and objective functions for community formation. The goal is to develop a common framework with consistent mathematical concepts, metrics and methods.

As an introduction to the appended articles of this work, I introduce general background in network science with some well-known examples and basic concepts in the field of modelling social networks. Three representative ideas are selected for this purpose: (i) Burt’s structural holes [1], (ii) Granovetter’s weak ties [2] and (iii) Centola’s ideas of complex contagion [3]. The concepts of complex contagion and processes on networks are closely related to this study. Structural holes and weak ties represent different and somewhat opposite views of analysing and interpreting social networks. These examples demonstrate that no one model fits all aspects of social networks. Different applications of social network analysis exist where these methods and models can be used to describe real-world phenomena.

The primary focus of this study is on spreading models applied to social interactions and epidemic spreading. Spreading models can be used also to describe information transfer in communication networks but not to consider infrastructure networks, like communication networks or power grids. However, I discuss differences in information and influence spreading in social networks. As an example, I show that a classical network connectivity model can be used, instead of the spreading model, as an alternative and different model to describe orchestrated processes between modules in a brain network.

This study consists of the following main parts: In Sections 1—3, I discuss general ideas of network science, the research objectives of this study and related works. In Section 4, I present a novel influence spreading model and its sub-models. In Section 5, I present some representative applications of the model. Finally, in Sections 6—8 I summarise the results, provide examples of future research and draw conclusions.
After the general introduction in Sections 1.1–1.3, I pose the research questions and delimitations in Sections 2.1–2.2 and discuss research gaps in Section 2.3. In Sections 3.1–3.5 I focus on presenting research methods and reviewing related literature. Related works are presented in the areas of network models, social influence analysis, centrality measures, community detection and epidemic models.

1.1 Network Science

Network Science is a growing field of research and it has gained a lot of interest in many application areas [4]. One definition of network science is the study of network representations of physical, biological and social phenomena aiming at predictive models of these phenomena. In addition to specialised methods of network modelling, networks can be studied with general methods of computer science, statistics and many other fields of mathematics. 

One narrow characterisation of network science is an academic field that studies complex networks considering distinct elements or actors represented by nodes (or vertices) and the connections between the elements or actors as links (or edges). We can say that any system having dependencies or relationships between its subsystems or elements can be modelled with a network structure or topology. Graph theory alone is a broad theoretical branch of mathematics.

Its applications cover areas such as telecommunication, computer networks, transportation, power grids, biological networks, trade networks, semantic networks, and social networks. In social networks, people are connected by profession, friendship and family ties. Studying social networks is increasingly important in order to understand how knowledge, behaviour and resources spread in society. Social media networks are a class of social networks that have become very popular with the increasing use of the Internet and the World Wide Web (WWW). The Internet is a physical network of computers linked by cables or radio links and WWW is a network of web pages and the links between them. Applications of web search exploit data mining technologies together with different algorithms based on network models. Also, one growing area of network science is the study of cellular networks where interactions between genes and proteins are described as network structures. In biochemistry, intermediate or end products of a set of life-sustaining chemical reactions (metabolites) can be described as network structures. Modelling the resilience of communication and infrastructure networks is an important practical application in the interdependent world of technological networks. [4]

Much of the current research of applied network science draws on diverse statistical, data analysis and data mining techniques. These methodologies can be used together with methods based on network structure, or independently, especially when the structure and properties of the investigated system are not fully known. In practice, this is a common state of affairs, because of insufficient data or the system itself is not fully understood. Examples of the two cases are partial or inaccurate information about relations and their strengths in social networks and the functioning of brain networks. Complex systems and complex
networks are two closely intertwined and related concepts reflecting the two overlapping views.

Complex networks and complex systems have been taken almost as synonyms but here the focus is mainly on complex networks. In many cases, complex networks have been modelled with basic topological terms using graph theoretic concepts like node degree, clique, the shortest path between node pairs and other local descriptions of a network. The degree of a node is the number of connections that it has to other nodes in the network. A clique is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent [4, 5]. The shortest path, or geodesic path, between two nodes in a graph is a path with the minimum number of links. If the graph is weighted, it is a path with the minimum sum of link weights. Studies based on these concepts provide insight into the understanding of networks and especially enable the comparison of research results in a standard way.

Model selection is important to gain the best possible results and, on the other hand, to avoid biased hypotheses. It is self-evident that unsuitable models can lead to erroneous research results. The downside of using possibly too simple models is that they can be inadequate for explaining the complex phenomena we are investigating. It was discussed in a conference (the 7th Conference of Complex Networks and their Applications) that using a particular model, even though the problem or the sub-problem is determined, is not always properly justified. In many cases, the phenomenon itself is poorly known, e.g. functioning of the brain, or processes are very complicated, e.g. interactions in social networks. In the case of social networks, several models have been proposed for describing different social interactions. Examples of these are the models of structural holes in Burt’s model [1], Granovetter’s weak ties [2] and Centola’s complex contagion interactions [3]. In addition, it is possible that the same models or their variants, possibly with different parameter values, can be useful for examining different network phenomena.

Despite the multitude of methods and models, many areas of network science research are still in a developing phase. Many types of processes and dynamics on networks are complex functionalities that have not been adequately described in current network models and methodologies. Social interrelationships in human social networks have many conceptual levels and one individual is a member of different social groups in family, work and free time relations. One consequence of this is the status of a research problem called ‘Community Detection’. Maybe hundreds of community detection methods have been proposed but no commonly accepted community detection method or a definition of the concept of a community exists. This has been criticised in several review articles discussing and comparing community detection methods and algorithms [6].

Social network analysis focuses on studying social relationships with the methods of network science. Social structures have been investigated through social network analysis including information circulation, friendship and acquaintance networks, social media networks, business relationships, and disease transmission. Social network analysis techniques are commonly used in sociology and it has become an important technique in modern sociology. It has
also gained a significant role in research areas of anthropology, demography, communication studies, economics, history, information science, organisational studies, political science, public health and social psychology. [4, 5]

Dynamics on networks have traditionally been modelled with simulations and random walks. Analytical formulas have been developed for specific network structures such as scale-free networks [5]. Complex topological structures make analytical solutions difficult or in many cases infeasible. Therefore, more research is needed in this area as the majority of applications of network science involve some kind of temporal processes on networks or changes in structures of networks. [7]

1.2 Modelling Social Networks

In this subsection, some concepts and models of network modelling are introduced. Burt’s structural holes [1] and Granovetter’s weak ties [2] are probably the best-known classical ideas in the context of social network modelling. Complex and simple contagion are modern concepts that are more closely related to this study. The motivation for discussing Burt’s and Granovetter’s work is to emphasise the existence of different social interactions and corresponding models. It can be concluded that one model does not fit all purposes and applications. After these different and complementary approaches, I will bring up some toughs about the processes on complex networks.

1.3 Basic Concepts of Network Modelling

Networks are modelled with nodes and links between nodes. Links describe connections, interactions or other relations between nodes. Links can be bi-directional with different strengths. Strengths can describe different aspects such as the strength of social ties or the capacity of communication lines. Much of the complexity comes from different ways of connecting nodes with links between them. Most real-world networks have a non-trivial topology. For example, in social networks, not all members of the network are connected and the strengths of connections are different.

Standard problems in analysing networks are to search important or central nodes in networks and discover communities or modules in network structures. Different definitions of what we mean by important, and using different models, can provide different results. The quantities like the degree, closeness centrality and betweenness centrality are the most common centrality measures [8]. Rankings of nodes and detected communities are different and depend on the model-specific measures of importance and coherence of communities. It is also possible that the same model can use different measures for quantifying the solutions from different viewpoints.

One definition of a community is a locally dense connected sub-graph in a network. Modularity maximisation and spectral graph partitioning are two examples in the wide context of community detection methods. Modularity measures
the strength of the division of a network into modules. Modularity can be defined [9] as the fraction of links falling within the given groups minus the expected fraction if links were distributed at random (The formal definition of modularity will be presented later in Eq. 12). [4, 5]

1.3.1 Structural Holes

Burt has introduced the concept of structural holes [1] in an attempt to explain the differences in social capital. A structural hole is a gap between two individuals with complementary resources or information. Burt’s theory suggests that individuals hold positional advantages or disadvantages from how they are embedded in social structures. Borgatti has expressed a simpler alternative to Burt’s original formulas in terms of redundancy [10]. The general meaning of redundancy is that a person’s ego network has redundancy to the extent that her contacts are connected as well. The redundancy of an egocentric network is

\[
r = \frac{2t}{n},
\]

where \( t \) is the number of connections in the network (not including connections to ego) and \( n \) is the number of neighbouring nodes (excluding ego). We can then define the ego’s effective size as:

\[
e = n - \frac{2t}{n}.
\]

A network structure used as an example in [1] is shown in Fig. 1. Node \( G \) is the ego in the network. From Eq. 2, we get for node \( G \) the value of effective size \( e_G = 6 - \frac{2 \times 4}{6} = \frac{4}{3} \). Similarly, for nodes \( A, B, C, D, E \) and \( F \) we have \( e_A = 2.5, \ e_B = 1 \\frac{2}{3}, \ e_C = 1, \ e_D = e_E = e_F = 1 \).

![Figure 1. Ego-centric network structure adapted from [1].](image)

<table>
<thead>
<tr>
<th>Node</th>
<th>Effective size</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2.5</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>1.67</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>D-F</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>G</td>
<td>4.67</td>
<td>6</td>
</tr>
</tbody>
</table>
1.3.2 Weak Ties

Granovetter’s article ‘The Strength of Weak Ties’ is a widely cited paper in social sciences [2]. It hypothesises that more novel information flows to individuals through weak rather than strong ties. The reason for this is that our close friends tend to move in the same circles that we do, and the information they receive overlaps considerably with what we already know. Acquaintances, by contrast, know people that we do not know and thus receive more novel information. In marketing, information science or politics, weak ties enable reaching populations and audiences that are not accessible via strong ties. [2]

Strong and weak ties have two meanings: relational at the dyadic level and structural at the population level. The first refers to the strength of the influence that is relayed through the tie. The structural strength of a tie refers to the ability of a tie to facilitate propagation by linking otherwise distant nodes in a social network. Granovetter’s insight is that ties that are weak in the relational sense are often strong in the structural sense because they provide shortcuts across the social network structure. The strength of weak ties is that casual relations are more likely to be formed between socially distant actors with few neighbours in common. Long ties greatly increase the rate at which influence propagates, despite the weakness of the tie as a conduit. [2, 11]

1.3.3 Complex Contagion

Complex contagion in social networks refers to how multiple sources of confirmation are required before an individual adopts a change of behaviour such as participating in a social movement or adopting an innovation [3, 12, 13]. Unlike simple contagion of disease or transmission of information, it may not be possible for the behaviour to spread after only one incident of interaction with a neighbour. There is a distinction between the acquisition of information and the decision to act on the information. The decision to act on the information is an active choice. In making that choice, people are influenced not only by the content of the information but they are also influenced by the observation of prior adopters, especially those that they know. The principal thesis in [3, 12] is that network structures that are highly efficient for the rapid dissemination of information are often not conducive to the diffusion of collective action based on the information.

The spread of influence through complex contagion in a social network may depend on many factors; for instance, how many of one’s friends adopt the new idea or how many of them cannot influence the individual and their tendency in adopting the change. Four mechanisms that explain why complex contagions require exposure to multiple sources of activation have been pointed out in [3, 12]: strategic complementarity, credibility, legitimacy and emotional exchange.

Complex contagion phenomena are commonly modelled as processes where fractions of adopting neighbours necessary for exposure are set as individual thresholds. This idea was first introduced by Granovetter [14] who discussed the ideal network structure and threshold distribution to allow for the evolution of riots or other collective movements. [15]
1.3.4 Processes on complex networks

In dynamic systems, the state of a system changes over time according to some given rules. Epidemic spreading in populations, influence spreading in social networks and the flow of traffic on roads are important practical applications. In some specific applications, it is not clear which kind of models best describe the system, or whether both static and dynamic behaviour can coexist in the same system. Standard approaches to study dynamical processes on networks rely on simulations because analytical mathematical expressions are not available or they are complicated. However, remarkable research has been conducted in percolation theory where analytical solutions exist for some network topologies. These methods are not directly applicable for modelling the detailed topology of empirical networks. [7]

By definition, processes on networks have a time-dependency. Two examples of processes on networks are spreading on network structure and changing network topology. Spreading processes start from a node or a set of nodes and propagate between nodes via links in the network structure. Also, the network structure itself may change during the spreading process. Changes in network structure and changes in link and node attribute values are common in many applications. For example, virus-spreading in computer networks or human social networks is controlled with virus protection software or vaccination programs in the human population.

Spreading processes on networks cover a variety of situations as processes can depend on the states of other nodes or links in the network. For example, virus spreading may not be possible or it is only partial when nodes are immunised after vaccination or recovering from an infection. Another example is that information or rumours are spread more actively when heard for the first time. In social systems, many overlapping processes, including both information and influence spreading, are concurrently changing our beliefs and opinions.
2. Research Objectives

The research methods of this work follow the usual conventions in natural sciences. I have reviewed the existing literature on social network models and tried to find research gaps in that area. After discovering research gaps I have aimed to build novel models that are extendable to several problems in modelling social networks and epidemic spreading. The new models have been validated by comparing their outcomes with results from earlier research of comparable studies or, if available, with real data. Comparing with the existing literature can also cover theoretical results or metrics from other models. Finally, I have focused on investigating models and analysing their representative numerical results.

2.1 Research Question

The primary research question of this study is the following:

*How to develop an influence spreading model that describes the detailed network structure of a network that considers all paths between nodes in the network?*

In the context of the influence spreading model, there are the following sub-questions:

1. *How can we define generalised centrality and betweenness measures in the model?*

2. *How can we detect communities and sub-communities in the model?*

3. *How can we define quality measures (objective functions) for the detected communities in the model?*

4. *How can we incorporate different temporal distributions into the model?*

5. *How to apply the model for describing epidemic spreading?*
2.2 Delimitations

In this study, I focus on analytical and probabilistic models and delimit simulation models outside the scope of this study. Social network analysis and epidemic spreading are the main application areas but I will also provide examples of how to apply the same models to other areas like communication networks and brain networks. I delimit pure sociology, epidemiology and biomedical sciences outside the scope of this work.

Many aspects require more and new research, for example, synthetic benchmarks and a thorough comparison of the proposed measures with the existing ones are left as future research (see Section 7). All the models and measures proposed in this study are based on probabilistic quantities. As the proposed measures have a probabilistic interpretation, and this is why numerical comparisons with other measures without probabilistic interpretation would not be very informative, I compare the proposed definitions of this study with other definitions that have some resemblance with my definitions. The proposed node centrality measures and the community detection method are based on the same probabilistic measure, the influence spreading matrix. This is why building a synthetic benchmark to compare the proposed community detection method with other community detection methods is also a subject for future investigation.

2.3 Research Gaps

In the current literature on modelling social networks, models are not adequate for describing spreading phenomena in the entire network structure. The reason for this is that theoretical analytical models often are limited to only local interactions [16] or specific network topologies such as scale-free networks [5]. In addition, limited computational resources have been restricting the use of simulation models. Models confined to local interactions do not consider global interactions, which is the property needed in many applications. For example, the effects of peripheral nodes or placements of sensors in a network topology require considering interactions among nodes with all distances between each other. [17, 18]

Many existing models have limited capabilities of including detailed features of node, link and network characteristics. Node and link weights and directionality are essential factors in calculating accurate centrality measures, community structures and analysing spreading processes between different parts of the network structure. Detailed network structures play a role in many applications like building the optimal formation of teams, and analysing individual nodes or groups in a network structure. Calculations of process quantities on the network structures with weighted and directional links provide more accurate results than the corresponding calculations based on the assumption that all links are bi-directional with the same strengths. Examples of processes on networks are the propagation of different kinds of influence like behaviour, opinion and belief. Also, the processes of information and epidemic spreading on a network structure depend on link weights and directionality.
Standard centrality and betweenness centrality measures in the literature have been defined independently of any network model. Maybe a more concise way of defining centrality and betweenness measures would be to use common concepts with a network model, for example, an influence spreading model. Then, all quantities of the model would be based on the same framework. In the spreading model, spreading probabilities, centrality and betweenness measures, and other derived metrics would be consistent. [8]

One characteristic of current propagation models using message passing methods for the solution of probabilistic models on networks such as epidemic models, spin models and Bayesian graphical models suffer from the shortcoming that they work poorly in the common case of networks that contain loops and repeated events [19]. Circular and repeated events are typical for influence spreading processes on social networks. Common to social interactions and many epidemics spreading incidents is that they spread through human contacts and, in particular, breakthrough events are possible. Breakthrough spreading can be modelled as repeated events through nodes in the spreading process on the network structure.
3. Related Work

In this section, I will go through some related research to introduce commonly used terms and concepts that are needed in subsequent sections. Related research is presented under the following sub-sections: ‘Social influence analysis’, ‘Centrality and betweenness measures’, ‘Network models and community formation’, ‘Community detection’ and ‘Epidemic models’. These sections introduce basic methods in the literature and serve as background for the models and applications in this study. How these topics and methods are related to the present work will come through in the subsequent sections ‘Influence Spreading Model’ and ‘Applications of the Model’ where corresponding concepts and methods are presented. Rather than covering completely the broad area of network modelling, I refer to the books written by Barabási [4] and Newman [5], where the general concepts and literature reviews of network science have been broadly presented. In the case of epidemic spreading models, I have selected, in addition to the classical SIS- and SIR-models [20, 21], some recently proposed models from the vast amount of literature in this field, in order to illuminate the differences they have with the present work.

3.1 Social Influence Analysis

State-of-the-art work in models, methods and evaluation aspects related to social influence analysis has been presented in [22]. The authors consider both microscopic and macroscopic level social influence models. As a future research direction, they suggest that macroscopic models should focus on how to consider human behaviour and different mechanisms during influence spreading processes.

Borgatti [23] has suggested a typology of flow processes based on two characterisations: the kinds of trajectories that traffic can follow and the method of spread. Specifically, in an influence process, individuals affect changes in each other’s beliefs or attitudes. The attitudes spread through replication rather than transfer where the trajectories followed by the attitude can revisit nodes. One can continue to influence somebody about the same thing over time. [23]

The threshold and cascade models for the diffusion of innovation and maximising the spread of influence through a social network have been studied in an article [24]. Recently, a long-standing problem of how to model belief propagation in systems with short loops has been solved [19]. Another study [25] demonstrates that diffusion probability along a social tie is a product of power-laws of degrees of the disseminator and the receiver. Here it is shown that the
information diffusion cascades generated by the model fit the structural properties of the observed systems’ cascades in three peer-to-peer online social platforms.

Complex contagion [12] is a phenomenon in social networks in which multiple sources of reinforcement are required before an individual adopts the change of behaviour. It differs from simple contagion in that it may not be possible for the influence to spread after only one incident of contact with an infected neighbour. The spread of complex contagion across a network of people may depend on many social factors, for instance, how many of one’s friends adopt the new idea. The authors suggest different mechanisms of complex contagion. Emotional contagion is one of the mechanisms that explain the need for multiple exposures in the spread of influence. Centola’s work builds on threshold models of collective behaviour. [3]

3.2 Centrality and Betweenness Measures

Nodes in a network can have central or peripheral positions in the structure. Degree centrality is the simplest closeness centrality measure. The degree of a node is defined as the number of nodes connected to it. This is a local measure and it does not take into account the position of the node in the network. Node’s closeness centrality measures how central or influential the node is to other nodes. On the other hand, betweenness centrality measures the role of a node as a proxy between other nodes in the network. The closeness and betweenness centrality measures have many variants and the choice of the specific measure depends on the application. Mathematical models provide other centrality measures such as the eigenvector centrality. One can also define out-centrality and in-centrality in directed networks. As for a node’s betweenness normally it is not defined for inward and outward directions.

From the review article [8], I have selected definitions and descriptions of eight centrality measures (Eqs. 3―10) that are closely related to the present work. In particular, I have selected measures that consider the structure of the whole network or are global in that sense. The classical closeness centrality measures in Eqs. 3―4 and the betweenness centrality measure in Eq. 5 are local measures.

The closeness centrality measure in Eq. 3 is based on the inverse sum of the shortest distances to the other nodes of the network. The distance in the network is defined as the minimum number of links needed to move from one node to another. This is the same as the length of the shortest path between the two nodes.

The *closeness centrality measure* in Eq. 3 is based on the inverse sum of the shortest distances to the other nodes of the network. The distance in the network is defined as the minimum number of links needed to move from one node to another. This is the same as the length of the shortest path between the two nodes.
The normalised version can be obtained by multiplying the formula by \((N - 1)\), where \(N\) is the number of nodes in the network. Here the factor \((N - 1)\) is justified because the distance from a node to itself is zero.

If the network is not connected, some of the distances are infinite and the closeness centrality of all nodes becomes zero. An alternative definition avoids this by defining another normalised version of the closeness centrality measure as

\[
C_i^{(\text{clos2})} = \frac{1}{N - 1} \sum_{j=1}^{N} d_{i,j}^{-1}.
\] (4)

The betweenness centrality measures the ability of a node to mediate influence between nodes in the network. The betweenness centrality of a node is based on counting how often it falls in the communication paths between pairs of nodes. Nodes having high betweenness centrality values can control the flow of information or influence through the network. The standard betweenness centrality measure is defined as

\[
C_i^{(\text{betw})} = \frac{1}{(N - 1)(N - 2)} \sum_{s,t=1}^{N} \frac{\sigma_{s,t}(i)}{\sigma_{s,t}}.
\] (5)

The sum is taken over all node pairs excluding node \(i\). The number of the shortest paths from source node \(s\) to target node \(t\) is denoted by \(\sigma_{s,t}\), and \(\sigma_{s,t}(i)\) denotes the number of paths that include node \(i\). The definition is valid for both directed and undirected networks and it measures the average fraction of shortest paths that cross a node.

Katz centrality generalises the degree and closeness centrality measures by taking into account not only the immediate neighbours or not only the shortest paths from a node to other nodes [26]. Katz centrality is defined as

\[
C_i^{(\text{Katz})} = \sum_{k=1}^{\infty} \sum_{j=1}^{N} \alpha^k (A^k)_{j,i},
\] (6)

where the power of matrix \(A^k\) accounts for the number of paths of length \(k\) between the pair of nodes. The adjacency matrix \(A = (A_{i,j})\) encodes the full topology of the network as \(A_{i,j} = 1\) if there is an edge from node \(i\) to node \(j\), and \(A_{i,j} = 0\) otherwise. A decay parameter \(\alpha < 1\) is introduced to weight the contributions of nodes at increasing path lengths.

The eigenvector centrality is a measure that depends recursively on the centralities of the node’s neighbours. Katz centrality, hubs and authorities centrality and PageRank are variants of eigenvector centrality [8]. The basic eigenvector centrality measure is used only for undirected networks whereas the three variants are also appropriate for directed networks. The hubs and authorities centrality (Eqs. 7–8) assigns to each node two different measures both for
sending and receiving influence. The definitions for incoming and outgoing
measures are

\[ C_i^{(auth)} = \alpha \sum_{j=1}^{N} A_{ij} C_j^{(hub)}, \quad (7) \]

\[ C_j^{(hub)} = \alpha \sum_{i=1}^{N} A_{ji} C_j^{(auth)}. \quad (8) \]

For undirected networks, this definition coincides with the basic eigenvector
centrality and the distinction between the hubs and authorities disappears. [8]

The PageRank has been used as a centrality measure in the Google search en-
gine. The PageRank centrality measure penalises contributions from nodes hav-
ing a large number of outgoing links. The PageRank centrality measure is de-

\[ C_i^{(PR)} = \alpha \sum_{j=1}^{N} A_{ij} \frac{C_j^{(PR)}}{k_j^{out}} + \frac{1 - \alpha}{N}, \quad (9) \]

where the parameter \( \alpha \) controls the relative importance of the eigenvector and constant terms, and the out-degree is denoted by \( k_j^{out} \).

The random walk centrality is defined by identifying the probability of a ran-
dom walker following a path from node \( i \) to node \( j \) as

\[ P_{i,j} = \frac{A_{i,j}}{k_i^{out}}. \quad (10) \]

Both the random walk closeness centrality and random walk betweenness cen-
trality based on the above matrix \( P \) have been defined in [27].

### 3.3 Network Models and Community Formation

In many complex systems, dependencies are modelled with networks consisting
of nodes and links [4, 5]. The nodes can have different roles as central influenc-
ers or mediators depending on their strengths and location in the network struc-
ture. Both the links and nodes can have attributes that describe their static and
dynamical properties. Examples of these are probabilities of functioning links
or nodes and probabilities of spreading influence over links or nodes. Commu-
nity detection methods have been developed to understand the formation of
groups in social networks or to categorise sub-systems in technological and bi-
ological networks. Community detection methods can be based on spreading
models as in the present work or static network models as in the classical net-
work connectivity model.

The classical network connectivity model is designed for describing the reliabil-
ity of communication networks [28]. If the reliability values between all the
neighbouring pairs of nodes in the network are known, the reliability values be-
tween any pairs of nodes in the network can be computed. Here, the reliability
is identified with the probability of a functioning connection. From the general
reliability theory [28] the reliability of a network \( V \) is defined as
\[ r(V) = \sum_{s \in \mathcal{S}} \prod_{e \in s} (1 - p_e) \prod_{e \notin \mathcal{S}} p_e, \]  

where \( S \) is a set of links where the network is connected and \( \mathcal{O} \) is the set of all connected states of the network. Links are denoted by \( e \) and the probability of a functioning link is denoted by \( p_e \). The above equations are polynomials of the order of the number of links \( N_L \) in the network. In this form, the equations describe the reliability of an entire network. In our case, I will apply the results for pairs of nodes by only taking the relevant terms in the summations.

There may exist hundreds of conference articles proposing different community detection methods on a local scale, or a microscopic level, using some phenomenological rule for deciding the community to which a node belongs. The network model is implicitly contained in the rule of community formation based on a node’s neighbourhood. Here, I do not review local network models because they are outside the scope of this work as the model can be characterised as global or macroscopic.

In local community detection methods, a network model can be understood both as the physical structure consisting of nodes and connections between them and also as local process characteristics of the network structure. However, in more general models the physical structure and process behaviour can be considered separately. In the next sub-section, some basic community detection methods are discussed.

### 3.4 Community Detection

Community detection is one of the most important applications of complex networks. A great number of different community detection methods and algorithms have been published in the literature [6]. Classical graph partitioning is the problem of dividing the nodes of a network into a given number of non-overlapping groups of given sizes such that the number of links between groups is minimised.

A problem with the various approaches is the lack of a commonly accepted definition of a community. On the other hand, one definition for all purposes probably is not possible because of different data available and different requirements in applications. In many cases the concept of a community is not defined explicitly, instead, the used method and algorithm define the concept implicitly. In this way, the number of different definitions of a community will increase with versions of new algorithms.

Most of the community detection methods are based on optimising a quality function. A basic choice for the quality function is modularity [9]

\[ M = \frac{1}{2m} \sum_{v,w} \left( A_{v,w} - \frac{k_v k_w}{2m} \right) s_v s_w + \frac{1}{2}, \]  

where \( v \) and \( w \) are nodes in the network, \( 2m \) is the number of stubs in the network, in which half edges after each edge are cut into two halves is called stubs. The node degree of node \( v \) is \( k_v \), \( A_{v,w} = 1 \) means that there is a link between
nodes $v$ and $w$, and $A_{v,w} = 0$ means that there is no link between the two nodes. The matrix $A$ is the same adjacency matrix as in Section 3.2. The membership variable $s_v$ indicates if node $v$ belongs to a community: $s_v = 1$ if node $v$ belongs to community 1 and $s_v = -1$ if node $v$ belongs to community 2. The formula for $M$ holds for partitioning into two modules but it can be generalised for partitioning into a desired number of modules. In matrix terms, Eq. 12 reads as follows

$$M = \frac{1}{2m} s^T B s,$$

(13)

where $B_{v,w} = A_{v,w} - \frac{1}{k_v k_w} (2m)^{-1}$ is an element of the modularity matrix. Equation 13 for $M$ is similar in form to an expression used in spectral partitioning of graphs for the cut size of a network in terms of the graph Laplacian. This similarity can be used to derive a spectral algorithm for community detection [5]. The eigenvector corresponding to the largest eigenvalue of the modularity matrix assigns nodes to communities according to the signs of the vector elements. [4]

The Louvain algorithm and Infomap are two fast algorithms for community detection. These algorithms have gained popularity because of their suitability for identifying communities in very large networks. Both algorithms optimise a quality function. For the Louvain algorithm the quality function is modularity and for Infomap an entropy-based measure. In the Louvain algorithm, the modularity is optimised by local changes of the modularity measure and communities are obtained by aggregating the modules for building larger communities. Infomap compresses the information about a random walker exploring the graph. [4]

Stochastic block models have been used as a method for detecting community structure in networks and also for generating synthetic benchmark networks [29]. Many community detection methods discover also hierarchical and overlapping sub-communities in complex networks [5, 30, 31]. In a recent study [32] an information-theoretic method has been presented for discovering groups, or building blocks, of network nodes that are usually found together in the same community.

### 3.5 Epidemic Spreading Models

Contact networks between humans or animals as well as many other networks are dynamic. Temporal networks are a modelling framework where the time of contact is explicitly taken into account. A class of time-varying networks is adaptive networks where edges or their weights vary over time. The dynamics of links or nodes in adaptive networks can be influenced by dynamics in networks such as epidemic spreading [7]. Particularly, in the area of dynamic complex networks, more research and new models for specific applications are needed. In the following sub-sections, I present the basics of classical SIS and SIR models and a short overview of some other epidemic spreading models, related to the present work, are presented.
3.5.1 SIS and SIR models

Relatively simple models of epidemic processes tell us how the infection spreads in a population often realistically enough in terms of quantities such as the epidemic threshold, speed of spreading and the fraction of infected individuals at the final phase of spreading.

Various epidemic spreading models have been presented in the literature [33, 34, 35]. Simple compartmental models were proposed in [20]. Homogeneous mixing was assumed in which individuals are assumed to interact homogeneously with each other at random. In compartmental models, the population is divided into different classes depending on the stage of the disease such as susceptible, infectious and recovered [21]. Susceptible individuals can contract the infection, infectious individuals have contracted the infection and are contagious and recovered individuals have recovered from the disease. A recovered individual can be immunised and no longer contract the disease.

Traditionally, epidemics on networks have been modelled with SIS (Susceptible-Infected-Susceptible) and SIR (Susceptible-Infected-Recovered) models [20, 21]. The SIS model assumes two states, susceptible and infected, for each node. When a susceptible node is in contact with an infected node, the susceptible node changes its state to an infected node at a rate \( \beta \). If a susceptible node is adjacent to \( k \) infected nodes, the transition rate is equal to \( k\beta \). An infected node recovers at rate \( \mu \) and enters the susceptible state. Nodes can be infected repeatedly in the SIS model. The model can be used for modelling emerging infectious diseases and also viral spreading in online media.

Next, we assume that network structure does not affect epidemic dynamics. In a network where each node is adjacent to every other node, the SIS dynamics is given by

\[
\frac{dl(t)}{dt} = \beta (1 - l(t)) l(t) - \mu l(t),
\]

where \( l(t) \) is the fraction of infected individuals at time \( t \). By setting \( dl(t)/dt = 0 \) we can solve the fraction of infected individuals at equilibrium as \( l^* = 1 - \mu/\beta \) or \( l^* = 0 \). We obtain \( l^* > 0 \) if and only if \( \beta/\mu > 1 \). Thus, we conclude that the epidemic threshold is \( \beta/\mu = 1 \). If the infection rate is larger than the recovery rate, the infection continues. Otherwise, the infection will die out. Taking into account the network structure in uncorrelated random networks with a degree distribution the epidemic threshold is \( \langle k \rangle / \langle k^2 \rangle \) where triangular brackets denote the mean value. [5]

The SIR model assumes three states: susceptible, infected and recovered. The difference to the SIS model is that when an infected node recovers at rate \( \mu \) it enters the recovered state never returning to the susceptible state. A recovered node will not infect others and will not be infected. In the SIR model, the final populations consist of susceptible and recovered nodes. Typically SIR model calculations are started from a single infected node or a small fraction of infected nodes while all the other nodes are in the susceptible state. The main interests are the number of recovered nodes and the speed of infection spreading. The SIR dynamics is given by
\[ \frac{dI(t)}{dt} = \lambda(1 - I(t) - R(t))I(t) - \mu I(t), \]
\[ \frac{dR(t)}{dt} = \mu I(t), \]

where \( R(t) \) is the fraction of recovered individuals. Transition rates are denoted by \( \lambda \) and \( \mu \). If and only if \( \frac{dI(t)}{dt} > 0 \) at \( t = 0 \) the number of infected nodes increases to a maximum value before recovering. Therefore, the epidemic threshold in the SIR model (Eq. 15) is given by \( \lambda \frac{dI(t)}{dt} \bigg|_{t=0} \) resulting in the same threshold as in the SIS model (Eq. 14). For uncorrelated random networks, the epidemic threshold is equal to \( \langle k \rangle (\langle k^2 \rangle - \langle k \rangle) \). [5]

### 3.5.2 Selected Epidemic Spreading Models Related to this Work

Later, in the research on epidemic spreading, the social structure was studied where people are classified according to demographic information like age and gender. Refined models are built on contact networks wherein the explicit network of social interactions among individuals determines the conceivable virus propagation paths. Multi-scale models consider sub-populations coupled by movements of individuals where homogeneous mixing is assumed on the lower scale. Here agent-based models can be used to simulate the movements and interactions of individuals on a fine-scale structure [36]. [7]

The spread of diseases across modular networks has been studied in the context of international air transportation [37]. The approach allows a scheme for the containment of epidemics based on deactivating the most central links in transmitting the disease. The method accounts for the first-order correlations between links, although according to the authors, it could be extended to higher orders by assuming a higher analytical and computational cost. The model has been used to analyse the SIS epidemic spreading dynamics of the worldwide air transportation network.

Recently, a model of disease spreading in a structural modular complex network has been proposed [38]. The model is based on equations that describe the disease at the final state using link percolation theory and generating function formalism. The results in [38] demonstrate that the proposed model can be described at a global scale as an infectious transmission process between communities with global infectious and recovery time distributions that depend on the internal structure of each community and the number of bridge nodes that connect communities.

The epidemic model by [39] based on Microscopic Markov Chain Approach (MMCA) was applied for the case of Spain and it successfully forecasted the observed incidence in each autonomous region in Spain. The authors derive a mathematical expression that allows them to anticipate the results of mobility restrictions and confinement on the spreading of Covid-19 in any specific region. The relationship between the reduction of mobility and confinement measures and its impact on the spreading of the epidemics has a highly nonlinear effect on the reduction of the epidemics. The model enables policymakers to implement the right amount of mobility restrictions needed to minimise the infectious virus spreading.
4. Influence Spreading Model

In this section, I introduce the influence spreading model proposed in this study together with its sub-models for centrality and betweenness measures, temporal spreading distributions and community detection. The centrality and betweenness measures and community detection algorithm are based on the output of the influence spreading model, and the model for temporal spreading distribution is a sub-module in the influence spreading algorithm (P I; P II).

4.1 Influence Spreading Model

The influence spreading model to be presented in this section is the core of this study. Community detection and analysis methods are based on an influence spreading matrix that is the output of the influence spreading model. Also, novel centrality and betweenness measures are based on the same influence spreading matrix. The influence spreading model together with various measures based on the matrix constitute a consistent framework for analysing complex network structures and processes. First, I present the general properties of the influence spreading model and after that, I describe the mathematical definition of the model.

4.1.1 Properties of the Model

Influence spreading models are designed for describing complex social interactions and propagation of opinions, beliefs and behaviour in social networks (P I; P II). These events spread via connections, or paths, between people. Information content can vary, and the ways of social influence are changing and developing during the influence spreading process. I allow repeated attempts of influence from a source node to target nodes via all alternative paths. Not all attempts result in opinion changes of individual nodes, and nodes can continue to influence their neighbouring nodes even if they don’t change their own opinion.

Features of the influence spreading model are weighted links and nodes, directed links, and the possibility of using different forms of temporal spreading distributions. All these parameters of the model have real-world interpretations. Link and node weights are interpreted as probabilities of forwarding influence between neighbouring nodes and over a node, respectively. Spreading probabilities between all the pairs of nodes in a structured network can be calculated from the node and link weighting values and the temporal distribution function.
An influence spreading matrix or a probability matrix is constructed to describe influence spreading from each source node to all other nodes in the network. The influence spreading matrix $C(s, t), s, t = 1, \ldots, N$ is a $N \times N$ matrix where $N$ is the number of nodes in the network. The influence spreading model has been published previously in (P I; P II; P VII).

The model is based on analytical expressions for influence spreading probabilities via different paths from a source node to target nodes in the network structure. In the basic version of the model, the rate of spreading is assumed to be independent of the state of the network and its elements. Allowing loops and assuming state-independent propagation are consistent properties of a model that describes social influence and breakthrough effects. A model with self-avoiding paths and state-dependent propagation of events can describe formal information transmission when nodes have a memory (P IV). To avoid double-counting effects, common paths at their beginning from one source node to a target node are taken into account by applying the formulas of mutually non-exclusive events from probability theory. Different paths are assumed to be independent if they join or cross during the spreading process.

In the model, an important property of the spreading process on a network structure is the possibility of loops. If repeated visits on a node are allowed, one node can be visited several times during the spreading process. Here it is assumed that social influence and the spreading of beliefs and opinions can be modelled with such repeated events and gradual influence spreading as a function of time. In the process of spreading information and news, loops would be less probable. In the algorithm, it is also possible to set the maximum number of visits $\nu = 1, \ldots, L_{\text{max}}$, where $L_{\text{max}}$ is the maximum path length of computing spreading probabilities. Using the parameter $L_{\text{max}}$ is necessary to limit computing time in the case when paths with loops are allowed.

### 4.1.2 Definition of the Model

The influence spreading model has been implemented as a computer program whose pseudo-algorithm has been published earlier (P I; A I). The algorithm considers all possible paths from a source node to a target node. For every path, the probability of propagation is calculated by multiplying all the link weights along the path, where each link weight $w_{kl}$ from node $k$ to node $l$ is interpreted as the probability of propagation via the link. Let us denote a path $i$ with $\mathcal{L}(i)$ in the list of paths $\mathcal{L}$, then $W_{\mathcal{L}(i)}$ is the product of link weights $w_{kl}$ on that path. Node weights can be handled similarly but for simplicity, one can assume that node weights are equal to one $w_n = 1$ for all nodes. Time dependency is included in the model by a factor $S_{\mathcal{L}(i)}(T)$ where $L(\mathcal{L}(i))$ is the path length of path $\mathcal{L}(i)$. The probability of spreading is then $P_{\mathcal{L}(i)} = W_{\mathcal{L}(i)}S_{\mathcal{L}(i)}(T)$. The time-dependent factor approaches $S_{\mathcal{L}(i)}(T) \to 1$ when time $T \to \infty$.

Next, I describe an iterative formula for calculating the spreading probabilities in a network by considering paths between a source node and a target node. Paths from source node $s$ to target node $t$ are combined pairwise in descending order of the length of the common path before the first branching of the two
paths. Two paths \( \mathcal{L}(i_1) \) and \( \mathcal{L}(i_2) \) are combined into path \( \mathcal{L}(i) \) with index \( i \) in list \( \mathcal{L} \). The probability value of \( P_{\mathcal{L}(i)} \) is calculated by the following formula:

\[
P_{\mathcal{L}(i)}(T) = P_{\mathcal{L}(i_1)}(T) + P_{\mathcal{L}(i_2)}(T) - \frac{P_{\mathcal{L}(i_1)}(T)P_{\mathcal{L}(i_2)}(T)}{W_{\mathcal{L}(i_1)\cap\mathcal{L}(i_2)}S_{\mathcal{L}(i_1)\cap\mathcal{L}(i_2)}(T)}.
\]

Equation 16 is deployed in \( N_L - 1 \) iterative steps in the algorithm of the computer program, where the number of different paths from the source node to the target node is denoted by \( N_L \). Indexes \( i_1 \) and \( i_2 \) are selected according to the rule of descending path length of the common path part of the two paths \( \mathcal{L}(i_1) \) and \( \mathcal{L}(i_2) \). The common part of the paths is denoted here by \( \mathcal{L}(i_1) \cap \mathcal{L}(i_2) \). All the original paths in the list are processed only once and combined paths are processed in the same way as the original paths in the network. The length of the combined path is the maximum of the two path lengths: \( \max(\mathcal{L}(\mathcal{L}(i_1)), \mathcal{L}(\mathcal{L}(i_2))) \). In the procedure, the path \( \mathcal{L}(i) \) is inserted and paths \( \mathcal{L}(i_1) \) and \( \mathcal{L}(i_2) \) are removed from list \( \mathcal{L} \).

The iteration starts with two paths having the longest common path length \( \mathcal{L}(\mathcal{L}(i_1) \cap \mathcal{L}(i_2)) \) of the path \( \mathcal{L}(i_1) \cap \mathcal{L}(i_2) \) with \( P_{\mathcal{L}(i_1)}(T) = W_{\mathcal{L}(i_1)}S_{\mathcal{L}(i_1)}(T) \) and \( P_{\mathcal{L}(i_2)}(T) = W_{\mathcal{L}(i_2)}S_{\mathcal{L}(i_2)}(T) \), where \( W_{\mathcal{L}(i_1)}, W_{\mathcal{L}(i_2)} \) and \( W_{\mathcal{L}(i_1)\cap\mathcal{L}(i_2)} \) are the products of all link weights along paths \( \mathcal{L}(i_1), \mathcal{L}(i_2), \) and \( \mathcal{L}(i_1) \cap \mathcal{L}(i_2), \) respectively. If there are more than two paths with the same common path length, these paths can be processed in any order. Survival functions \( S_{\mathcal{L}(\mathcal{L}(i))}(T) \) describe the time-dependency of the probability of spreading over \( L(\mathcal{L}(i)), i = 1, ..., N_L \), links on paths \( \mathcal{L}(i) \) starting from a source node \( s \) with probability one. The number of different paths starting from the node \( s \) with path lengths \( L \leq L_{\text{max}} \) is denoted by \( N_L \). Mathematically, the survival function \( S_{\mathcal{L}(\mathcal{L}(i))}(T) \) can be expressed as \( 1 - F_{\mathcal{L}(\mathcal{L}(i))}(T) \) where \( F_{\mathcal{L}(\mathcal{L}(i))}(T) \) is the distribution function of the temporal spreading probability.

The probability of influence spreading between the two nodes is the final result of the algorithm after all paths have been processed. Thus for source node \( s \) and target node \( t \) we have

\[
C(s, t)(T) = P_{\mathcal{L}(N_L-1)}(T).
\]

Equation 17 defines the matrix element \( C(s, t)(T) \) of the influence spreading matrix \( C \) at time \( T \).

Two different forms of the survival function \( S_{\mathcal{L}(\mathcal{L}(i))}(T) \) have been discussed in (P II). The Poisson distribution describes a random response time and the e-mail forwarding distribution describes a process of receiving and forwarding messages. For the Poisson distribution, the time-dependent factor is

\[
S_{\mathcal{L}}(T) = 1 - \sum_{z=0}^{L-1} e^{-\lambda T} \frac{z^x}{z!}, S_0 = 1.
\]

The survival function of the e-mail forwarding distribution is calculated from the algorithm provided in the appendix of (P II).

Additional information about the algorithms and efficiency of their implementations can be found in (A I). Three pseudo algorithms in (A I) for the path combination process, complex contagion model and simple contagion model may
help understand the influence spreading model of this study. The complex contagion algorithm is a version for non-self-avoiding paths in the model and the simple contagion algorithm is the corresponding version for self-avoiding paths. (A I)

4.2 Scalability of the Influence Spreading Model

Alternative pseudo-algorithms and implementations have been published in (P I; P V; P VII; A I; A II). Specifically, a scalable algorithm is presented in (P I; A I) for the CC model where loops are allowed (non-self-avoiding paths) in the spreading process. An advantage of the model is its performance as networks up to about 100,000 nodes can be studied with the model. Even larger networks may be studied by limiting computations to shorter path lengths or a sub-set of the most influential nodes in the network. The key insight is that, with this model, spreading depends only on the distance from the source node, not on which path was traversed from the source node. This allows for a great improvement in time complexity, and the algorithm presented in (A I) can solve all pairwise spreading probabilities in \( O(N \cdot E \cdot L) \) time (where \( N \) denotes the number of nodes in the network, \( E \) the number of edges and \( L \) the path length).

Running times of the CC model program for selected real-world networks are listed in Table 2.

<table>
<thead>
<tr>
<th>Network</th>
<th>#Nodes</th>
<th>#Edges</th>
<th>Running time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArXiv collaboration network (ca-GrQc)</td>
<td>5,242</td>
<td>14,496</td>
<td>29.2 s</td>
</tr>
<tr>
<td>Facebook network (ego-Facebook)</td>
<td>4,039</td>
<td>88,234</td>
<td>37.5 s</td>
</tr>
<tr>
<td>Twitter network (ego-Twitter)</td>
<td>81,306</td>
<td>1,768,149</td>
<td>3.5 h</td>
</tr>
<tr>
<td>Google+ network (ego-Gplus)</td>
<td>107,614</td>
<td>13,673,453</td>
<td>50.7 h</td>
</tr>
</tbody>
</table>

The pseudo-algorithm presented in (P V; P VII; A I) illustrates my original idea (A III) for both cases of self-avoiding and non-self-avoiding paths. The implementation for self-avoiding paths is called Alg-ccsa in (P IV). Because it is based on keeping all possible paths from a source node to a target node in memory, it is not practical for larger networks, except with short path lengths, as it is known that, in the worst case, there are \( O(N^L) \) paths of length \( L \) in a network of \( N \) nodes. Self-avoiding spreading may be computed more efficiently using simulation as described in (P IV).

To summarize, the solution in (P I; A I) avoids the problem of keeping all paths in memory but it does not apply to the modelling of self-avoiding paths. On the other hand, the solution in (P V; P VII) is suitable for models using self-avoiding paths for moderate network sizes and even for large networks using short path lengths.

Numerical examples of computing times are provided in (P IV; A I). The results show that some versions of the proposed methods using self-avoiding
paths can be computationally heavier than direct simulation (See, for example, Alg-ccsa and Sim-sc in Table 3 in (P IV)).

Note, however, that there is a subtle difference between the Alg-ccsa and Sim-sc models. Alg-ccsa is based on analytical formulas for self-avoiding paths with breakthrough influence, while Sim-sc is a simulation model for self-avoiding paths without breakthrough influence. I am not aware of any analytical model corresponding to the Sim-sc model where breakthrough influence is not allowed.

In summary, I can say that the analytical and simulation methods may be described as dualistic approaches when their application areas and computation efficiencies are considered. (A I)

4.3 Centrality and Betweenness Measures

In this section, I present the definitions of out-centrality, in-centrality and betweenness centrality measures of this study. After that, I compare the definitions to other definitions proposed in the literature.

4.3.1 Definitions Based on the Influence Spreading Matrix

With the help of the influence spreading matrix $C$ defined in Eq. 17 out-centrality and in-centrality measures for nodes $s$ and $t$ respectively can be defined as

$$ C_s^{(out)} = \sum_{t=1}^{N} C(s, t), \quad (19) $$

$$ C_t^{(in)} = \sum_{s=1}^{N} C(s, t). \quad (20) $$

For simplicity, we do not indicate the time variable $T$ in Eqs. 19–23. Normalised versions of these centrality measures can be obtained by dividing the expressions by $N$ or $N - 1$ depending on whether the diagonal elements of the influence matrix $C$ are set to one or zero. Betweenness centrality measure for node $n$ can be defined as

$$ b_n = \frac{C - B_n}{C}, \quad (21) $$

where

$$ C = \sum_{s,t=1}^{N} C(s, t) \quad (22) $$

and $B_n$ is calculated similarly to $C$ with the node $n$ removed from the nodes of network $V$ in the network as

$$ B_n = \sum_{s,t=1}^{N, n\notin V} C(s, t). \quad (23) $$
When calculating the betweenness measure values, the normalisation with \( N - 1 \) instead of \( N \) together with diagonal elements of matrix \( C \) set to zero is preferred because this gives the intuitive betweenness value of zero for the nodes that are not connected to others.

### 4.3.2 Comparison with Other Centrality Measures

Definitions in Eqs. (19–23) are based on the influence spreading matrix \( C \) and have interpretations, depending on the normalisation, as probabilities or as numbers of influenced nodes in the network. In Section 3.2, I have summarised the most important centrality and betweenness measures used in the literature. These measures lack clear physical or probabilistic interpretation. For this reason, the mathematical expressions are provided for comparison and numerical results are kept to a minimum.

Closeness centrality in Eq. (3) and normalised closeness centrality in Eq. (4) are defined by simple phenomenological formulas based on the distances of nodes in the network, where the distance between a node pair is measured by the minimal number of links between the two nodes. Also, the betweenness centrality measure in Eq. (5) is based on the shortest paths between node pairs in the network. These measures are different from the proposed measures in Eqs. (19–23) in many ways. They don’t account for the node or link weights, they don’t allow modelling circular or self-avoiding interactions and they don’t enable modelling interactions as a function of time or path length. Generally speaking, definitions in Eqs. (19–23) don’t enable the modelling of different social or physical processes on network structures.

The proposed model of this study considers all the paths of length \( L \) in the network that are shorter or equal length than the maximum path length parameter value \( L_{\text{max}} \) (\( L \leq L_{\text{max}} \)). This kind of approach is called global instead of local [16]. The geodesic-based centrality measure proposed in [16] has been compared by a numerical example with the out-centrality measure of Eq. (19) in [P I]. The model in [16] considers all paths but it does not handle the combining process of paths that is unique to this work (See Section 4.1 or Algorithm 1 in [A I]). The models in [16] and our work provide similar results but there are some differences. For example, the model of this study, in the case where loops are allowed, shows that peripheral nodes in a network can have more spreading power than in the model [16]. This issue has been discussed in more detail and related references have been provided in [P I].

Katz centrality, Hubs and authorities centrality and PageRank centrality measures are defined recursively on the centralities of the node’s neighbours. They are variants of eigenvector centrality [8] and they are based on adjacency matrixes. These three centrality measures can be used also for directed networks. They are more related to the definitions in Eqs. (19–23) because they consider the entire network structure although they still lack a clear probabilistic or physical interpretation. The form of the Katz centrality measure resembles the definition of the out-centrality measure in Eq. (19) but the model does not have, just like the model in [16], a process for combining dependent paths in the network structure.
In Section 1.3, I have presented a couple of simple measures used in the literature to introduce basic concepts of network modelling. Burt’s model [1] serves as an example of a different approach and interpretation of social interactions.

In the next section, I discuss and compare the centrality and betweenness measures of this study in a small ego-centric network. Results from the Burt’s model and results based on the degree of a node and link weights are also displayed for comparison.

4.4 Comparison of the Centrality and Betweenness Measures in a Small Ego-centric Network

Next, I illustrate the centrality measures defined in the previous section with a simple example. The values of out-centrality, in-centrality and betweenness centrality measures of the ego-centric network are provided in Fig. 1. I discuss different network spreading models based on two properties of flow processes, state-dependency and the possibility of loops and refer to the concept of a node’s state when considering the probability of spreading through the node. If loops are allowed, one node can be visited several times during the spreading process. If loops are not allowed, only the self-avoiding paths are possible. The case of state-dependent processes and self-avoiding paths are categorized as simple contagion (SC) processes. All other alternatives are considered complex contagion (CC) processes.

Fig. 2 shows that for the low link weight value $w_l = 0.05$, out-centrality, in-centrality for the complex contagion (CC) model and centrality values for the simple contagion (SC) model are almost equal. Notice that in the SC model, out-centrality and in-centrality values are equal for undirected links with similar weights. This is why we have only one centrality value, and not out- and in-centrality values for the SC model. For the CC model, out-centrality and in-centrality values are different, except in the case of symmetric complete graphs. In addition, I show the results for Burt’s effective centrality [10, 1] and the results of the influence spreading model (P I; P II; P VII) for the limited path length values $L = 1$. Burt’s model takes into account redundancy (Eq. 1) in a person’s ego network when her contacts are connected as well (see the section ‘Structural Holes’). The values in Fig. 2 show generalised results of Burt’s effective size (Eq. 2) where link weights have been used in Eq. 2. For example, for node $A$ in Fig. 1, the effective centrality value is $0.05 \times (4 - 2 \times 3/4) = 0.125$. The results for $L = 1$ are proportional to the node’s degree because they are calculated as $w_l \times degree$. For example, for node $A$ the value is $0.05 \times 4 = 0.2$. 

27
Figure 2. Five alternative centrality measure values for the seven nodes \{A, B, C, D, E, F and G\} of the ego-centric network in Fig. 1. Results are shown for link weight values \(w_l = 0.05\) (top) and \(w_l = 0.5\) (bottom). Also, the average values of the five measures are shown. Out-C, CC and In-C, CC are the out-centrality (Eq. 19) and in-centrality (Eq. 20) values when loops and breakthrough effects are allowed during the spreading of influence on the network. C-SC show the in- and out-centrality values for self-avoiding paths and \(L = 1\) show the centrality values with the limited path length. Notice that \(L = 1\) values are proportional to the degree values of nodes. For comparison, I show also the effective centrality values of Burt’s model [1].

Next, the betweenness centrality measure [41, 42, 43] of Eq. 21 is investigated. Figure 3 shows the betweenness centrality values of the seven nodes of the ego-centric network of Fig. 1 as a function of link weights \(w_l\). The results are shown for the CC model (top) and the SC model (bottom). As expected, the ego of the ego-centric network, node \(G\) has the highest betweenness value. In both models, the betweenness values have a maximum value except for node \(C\) whose betweenness value is increasing as a function of the link weight value. In addition, in the SC model nodes \(D, E\) and \(F\) have increasing values as a function of link weight value. All nodes have an endpoint betweenness value when \(w_l = 1\), which
is the same for both models. From Eq. 22 we get the value $C = 49$. Because for node $G$ the value in Eq. 23 of $B_G = 27$, we get $b_G = (49 - 27) / 49 = 0.449$. And because for all the other nodes $n = A, ..., F$ we have $B_n = 37$, we get $b_n = (49 - 37) / 49 = 0.245$.

As can be seen in Fig. 1, node $G$ is the only bridging node between node $C$ and the rest of the network. Removing node $G$ divides the original network into two isolated parts. As expected, node $G$ has a high betweenness value because of its role as an important mediator in the network structure. The standard betweenness centrality measure of Eq. 5 has similar behaviour, although the numerical values of the two measures are not the same because Eq. 21 considers all possible interactions and paths in the network model. This is an important distinction to make as a lot of the structure of the graph is left out when we consider only the edges on the shortest paths from a node, which is the case with the standard definition.

![Figure 3. Betweenness values for the seven nodes (A, B, C, D, E, F and G) of the ego-centric network of Fig. 1 as a function of the link weight value $w_l$. The top figure shows the results when loops are allowed and the bottom figure shows the results for self-avoiding paths.](image-url)
Furthermore, the definition in Eq. 21 is more general in the sense that it is independent of the network model. Other models than the influence spreading model of Section 4.1 can be used as a basis in the definition as long as the influence spreading matrices, $\mathcal{C}$ in Eq. 22 and $B_n$ in Eq. 23, can be calculated from the network models. For example, the classical network connectivity model of Section 3.3 [28] can be used instead of the influence spreading model of Section 4.1. The proposed definition in Eq. 21 enables comparing different models of spreading or transmission and investigating their suitability for describing real-world processes on network structures. Figure 3 shows an example where betweenness centrality measures are calculated as a function of the link weight values $w_l$ for the complex (CC) and simple (SC) contagion modes as defined in this study.

Table 4 shows numerical values of the influence spreading matrix (Eq. 17) computed from three alternative spreading models for the network structure in Fig. 4. The three panels from top to bottom in Table 4 show the results of the Alg-cc, Alg-sacc and Sim-sc algorithms, respectively. The two horizontal panels show the results for two different link weights values $w_l = 0.05$ (left) and $w_l = 0.5$ (right). The content of Table 4 is summarised in Table 3. Interactions between the nodes of the network structure in Fig. 4 are described from nodes 1, 2, 3 and 4 as lines to nodes 1, 2, 3 and 4 as columns in the six matrices of Table 4. Line and column sums are also provided for the six matrices indicated by the orange, yellow and green colours. As a convention diagonal elements of the matrices are set equal to one: $C(i, i) = 1, \ i = 1, ... N$, for the network with four nodes $N = 4$.

Figure 4. This simple network structure is used to illustrate three alternative spreading models in Table 4.

Nodes 1 and 4 are in exchangeable positions in the network structure and similarly nodes 2 and 3 are in exchangeable positions if links are bidirectional with equal weights. This is reflected in the six matrices of Table 4. Minor differences between lines 1 and 4 or lines 2 and 3 in the Sim-sc algorithm (bottom panel in Table 4) are due to simulation errors. In small and sparse networks the three models provide almost similar results for low link values due to the converging of probability values along the paths in the network as can be seen from the
results for $w_l = 0.05$ in Table 4. However, for high link values, the three models provide different results as can be seen from the results for $w_l = 0.5$ in Table 4.

More interesting, than the trivial observations in the previous paragraph, is that influence spreading matrices are symmetric in the models where loops are not allowed as can be seen from the results for Alg-sacc and Sim-sc in Table 4. The results for the Alg-cc algorithm are not symmetric meaning that the probabilities can be different for the two directions of spreading between two nodes in the network. Mathematically this is expressed as $C(i, j) \neq C(j, i), i, j = 1, \ldots, N$. The effects of loops are more pronounced for high link weights as can be seen from the results for $w_l = 0.5$ in Table 4. For example, in the Alg-cc model $C(1,2) = 0.6969$ and $C(2,1) = 0.7946$, which can be explained by the fact that there is only one loop around node 1 and two loops around node 2.

Table 3. This table is a key to the structure of Table 4 and the content of the six matrices describing three algorithms and two values of the link weights.

<table>
<thead>
<tr>
<th>Alg-cc, $w_l = 0.05$</th>
<th>Alg-cc, $w_l = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alg-sacc, $w_l = 0.05$</td>
<td>Alg-sacc, $w_l = 0.5$</td>
</tr>
<tr>
<td>Sim-sc, $w_l = 0.05$</td>
<td>Sim-sc, $w_l = 0.5$</td>
</tr>
</tbody>
</table>

Table 4. Numerical values from the three spreading models of the Alg-cc, Alg-sacc and Sim-sc algorithms on panes from top to bottom respectively. Link weight values $w_l = 0.05$ (left) and $w_l = 0.5$ (right) are used in the computations.

4.5 Community Detection

The community detection method of this study is based on searching local maxima of the community influence measure of Eq. 24 computed from the influence spreading matrix elements $C(s,t), s,t = 1,\ldots,N$. The number of nodes in the network is denoted by $N$. Our model has the following form of the objective function for community detection:

$$P = \sum_{s,t \in V} C(s,t)(T) + \sum_{s,t \in (G-V)} C(s,t)(T).$$

(24)
The first summation is over the pairs of nodes \( s \) and \( t \) in a subset of the nodes in the network \( V \). The second summation is over the pairs of nodes in the remaining partition of the network denoted by \((G - V)\). Rows and columns are included in the sums that correspond to node pairs in the community, and node pairs that are not included in the community. Our approach allows detailed modelling of the networked system because all different paths between two nodes in a network are included in the model. The influence spreading model allows also loops, which can be restricted by allowing a limited number of visits, e.g. just one visit, on a node in the network. The maximum length \( L_{max} \) of paths in computing the values of \( C(s, t)(T) \) is necessary in the case of non-self-avoiding paths. The parameter is also useful with self-avoiding paths in investigating the effects of path lengths on the spreading process or in limiting computing times. The number of possible paths with self-avoiding paths with \( L_{max} = \infty \) is finite but in large networks, the number of different paths is very high.

The form of Eq. 24 is designed for processes allowing analysis at different time horizons during the spreading process. In addition, the method can be used for examining static network structures with the time of spreading approaching infinity \( T \to \infty \). In the model, individual weights for each node and link in the network can be used. As a consequence, the method enables studying the network with different interactions between the nodes. If such information is available, more exact results can be obtained.

The optimisation method assumes that the network is divided into two communities. However, the model provides multiple solutions for the local maxima of Eq. 24. Communities with high rankings according to the value of the quality function \( P \) in Eq. 24 are candidates for the split of the original community in real life. This may not be the most probable solution to the community formation process. I have also proposed a statistical measure for describing the probability of community formation (P II).

Matrix \( C \), which is used in the definition of the quality function is the influence spreading matrix, whose elements capture the influence of nodes on one another. For static networks, it is possible to use a different network model. As an example, I have used the network connectivity model (P II) to produce a time-independent connectivity matrix to be used instead of matrix \( C \) in Eq. 24.

Finally, I comment on the validation of the community detection method. Comparisons to real-world communities are provided in (P I; P II; P III). There are only a few ground truth community structures where the community has split into two or more divisions. In this category, the community detection model of this study has been compared with the classical examples of the Karate Club network and the animal social network of 62 bottlenose dolphins (P I; P II).

The model has been used to analyse real-world community structures also in cases where no actual splitting of the community has occurred. These kinds of examples are the Al Qaeda terrorist network (P III) and the Dutch students’ social network (P I). Examples of synthetic network structures are the game of Risk (P I) and the Les Misérables network (P II). Comparing different models with the synthetic network structures is only theoretical because no empirical
data on social interactions between the members of synthetic communities exist.

Figure 5. The results calculated with the community analysis method of this study show 11 strong divisions of the Les Misérables network. The borders of the divisions show building blocks of the community structure (P II).

Figure 6. Building blocks of the Les Misérables network were discovered in [32].

The Les Misérables network is chosen as an example because a research article investigating groups of network nodes (building blocks) in the Les Misérables network has been published recently in [32]. The study uses an information-theoretic method to discover sub-communities that are usually found together in the same community. The set of building blocks discovered in [32] can be
compared with the corresponding results of the model proposed in this study. The example network has a complex structure from which a set of sub-communities can be detected thus making it possible to compare different models. As can be seen from Figs. 5–6 where the information-theoretical results [32] and the results of this study (P II) are shown that the results agree with each other surprisingly well. The investigation of the Les Misérables network in [32] is important because the research method uses general information theory and it is not based on a particular community detection method proposed in the literature [6].

In Section 3.4, I have mentioned examples of other popular community detection methods. The Louvain algorithm and Infomap methods can be used to detect communities in larger networks than the method of this study. However, there is plenty of room for optimising the algorithm for searching the local maxima of Eq. (24).
5. Applications of the Influence Spreading Model

5.1 Social Networks – a Case Study

As a case study, I will focus my attention to model Al Qaeda’s social network structure before the tragic events of 9/11/2001. Till now social network analysis has provided a number of analytical methods to help in preventing or investigating terrorist and criminal activity. The two aspects of prevention and prosecution have been discussed in more detail in [44, 45, 46].

Here, I will present methods based on a modelling framework to assess the influence of a node in a complex network concerning influence spreading via different paths between the source node and target node. Different measures, such as closeness centrality, betweenness centrality, and community detection are computed with the same modelling framework. The proposed algorithm for detecting sub-communities uses influence measures, in which the influence of a node is considered in both directions: a node has influence on other nodes and a node is influenced by other nodes. Here the network topology and unequal weights of nodes and links are essential features of the model.

Next, I demonstrate the possibilities of discovering new phenomena and the usefulness of the method with real-world social network data. The terrorist network data of 19 hijackers have been collected from public sources after the terrorist attacks of September 11, 2001 [44, 45]. The aim of studying covert, illegal and terrorist networks is to uncover common and specific characteristics of these networks.

In [44, 45] the observations of the study are such that many of the hijackers were distant from each other in the network structure. In addition, many hijackers on the same flight were more than two links away from each other. Keeping cell members distant from each other and other cells minimises the possible damage to the network if a cell member is captured or otherwise compromised. There was a balance between security and resilience. Furthermore, trusted prior contacts were a strength in coordinating tasks. The 19 hijackers had other accomplices that were conduits for money and also provided needed skills and knowledge. Members of the sub-communities had different roles in planning, financing and carrying out the terrorist attacks. Sub-communities having only supporting functions were loosely connected to the more tightly connected core structures of the network.
The terrorist network has 62 nodes and 306 links between the nodes. The 62 members of the network were connected in many ways: attended the same college, took flight classes together, bought flight tickets using the same address, bought flight tickets together, were known to be together the week before the attacks and had last known address in the same area. Krebs [44, 45] determined three categories for the strength of ties between the members of the network.

Figure 7 depicts the network topology (structure) of the terrorist network and the results of the community detection algorithm (P I; P II) are indicated therein. The community detection algorithm detects seven sub-communities inside the terrorist network. These sub-communities are denoted by letters from A to G. I argue that the sub-community structures similar to those of Fig. 7 are characteristic of covert social networks. The sub-community D has a specific structure of central nodes and the sub-group G is inside this sub-community. Sub-communities A and B are connected to the core structures of sub-community D only with a few links. The members of sub-communities A and B did not get on the planes while the members of sub-community E were connected with more links to sub-community D due to their actual roles in the hijackings. The network topology of connections between the terrorists is partly self-organised, but the main operational principles were planned and implemented systematically before the attacks [44, 45].

The first aeroplane crashed to WTC north with five terrorists \{1, 9, 11, 12, 37\}. Node 1 is in a central role in the whole operation and is a member of sub-community D and the other nodes \{9, 11, 12, 37\} were members of subgroup G. The terrorists \{7, 8, 25, 26, 36\} crashed to Pentagon, and they were all members of sub-community E, while the terrorists \{2, 10, 13, 30, 34\} crashed WTC south, and they were members of sub-communities D and E with nodes \{13, 34\} in subgroup G. The four terrorists \{3, 29, 31, 38\} crashed in Pennsylvania, and they were members of sub-communities D and E with the node \{38\} also a member of subgroup G. All the nodes in subgroup G were hijackers on different planes such that they can be interpreted as having been the inner circle among hijackers.

Next, I will show some results of closeness centrality (Fig. 8) and betweenness (Fig. 9) measures with simplified parameter values when compared to the original article. In Figs. 8, 9 and 10 I use the node weight values of \(w_n = 0.05\) and set the time of spreading \(T = \infty\) (These parameter values are also used in other applications later in this study). Link weights \(w_l = 1.0, 0.75\) and \(0.5\) are used to describe the three categories of activity [44, 45] between members of the group. In Fig. 10, link weights \(w_l = 0.75\) are used to describe equal strengths of interactions between all pairs in the network structure. The results I show of the original paper and the simplified version are similar and this shows that the model is not very sensitive to the chosen parameter values. The time parameter \(T\) describes the temporal development of the spreading process (See Figs. 1–6 in P III) and low link weights with \(T = \infty\) (See Figs. 8–9) describe the final state of the spreading process. The results of these two calculations are close to each other but not precisely the same.
Figure 7. Structure of the terrorist network [44, 45] with detected sub-communities in the influence spreading model.

Figure 8 shows the closeness centrality values of Eq. 19 and Fig. 9 shows the betweenness centrality values of Eq. 21 for the 62 nodes in the terrorist network. I have indicated the 19 hijackers with lines on the horizontal axis to help us notice whether there are characteristic patterns in this group when compared with other nodes.

In Figs. 8—9, no clear patterns in the hijacker group are observed, but in a deeper analysis of Fig. 10, I can make some conclusions. In Fig. 10 the ratios of weighted results from Figs. 8—9 and unweighted results with $w_l = 0.75$ for all nodes are shown. From the ratios, one can easily recognise 18 individuals who have active connectivity in the network structure {3, 4, 5, 6, 8, 21, 22, 23, 26, 27, 28, 29, 30, 31, 34, 35, 53, 60}. Notice that, the most central nodes 1 and 2 are not in this set, but these nodes can be found both in closeness ($C_s^{(out)}$ of Eq. 19) and betweenness centrality ($b_n$ of Eq. 21) results. I listed nodes with their ratios above the threshold of the closeness $r_c \approx 1.01$ and betweenness $r_b \approx 1.1$ for centrality values $C_s^{(out)} > r_c$ and $b_n > r_b$.

The 18 active nodes form a sub-network of Fig. 1 but only in eight cases, a hijacker is in the group of active nodes. Nodes {3, 8, 26, 29, 30, 31, 34, 38} appear in both sub-groups of hijackers and active members (See Fig. 10). Ten members of the terrorists out of the 18 active members are not actual hijackers. This can be explained by psychological and behavioural factors or the operation was planned with different roles in mind.
Figure 8. The values of closeness centrality measure (Eq. 19) of the nodes in the terrorist network in Fig. 7.

Figure 9. The values of betweenness centrality measure (Eq. 21) of the nodes in the terrorist network in Fig. 7.

Figure 10. The ratios of the results with the weighted and equal weights are calculated from Figs. 8—9. Closeness centrality is shown by bars (left vertical axis) and betweenness centrality with dots (right vertical axis).

Values of link betweenness centrality are computed analogously to the node betweenness centrality in Eq. 21 where instead of removing a node, one link is removed from the network. Figure 11 shows the results of the 50 most important links \{1–2, 2–1, 1–39, 39–1, …\} of the terrorist network in Fig. 7. The results demonstrate the central role of node 1. Connections from nodes 2 and 39 to node 1 are the most important. The order of connections can be used for analysing the social network, or specifically, for prioritising the order of sending or forwarding information to optimise the influence in the network.
5.2 Social Media Network – an Example

Complex contagion is the phenomenon in social networks, in which multiple sources of reinforcement are required before an individual adopts the change of behaviour. It differs from simple contagion in that it may not be possible for the influence to spread after only one incident of contact with an infected neighbour. The spread of complex contagion across a network of people may depend on many social factors, for instance, how many of one’s friends adopt the new idea. The authors in [3, 12] suggest different mechanisms of complex contagion. Emotional contagion is one of the mechanisms that explain the need for multiple exposures in the spread of influence. Centola’s work builds on threshold models of collective behaviour.

I have studied different spreading models in the context of influence and information spreading on social networks (P IV). In the simple contagion (SC) models, only self-avoiding paths are allowed and nodes transfer influence to their neighbouring nodes only after their state changes. In this study, only the case of self-avoiding paths and state-dependent spreading is regarded as an SC process. All other alternatives are categorised as complex contagion (CC) processes. In the present approach, it’s typical for node pairs to repeat interactions in the complex contagion models. (P IV)

The example Facebook network in Fig. 12 shows some special properties of social media networks. Here it is discovered that the structure of this particular Facebook network shows high spreading probabilities in the Alg-cc curves in Fig. 13 at high maximum path lengths. These effects are due to highly connected clusters (See Fig. 12) that enable loops inside the structure. I conclude this from the fact that the Sim-sc results in Fig. 13 do not show similar effects. These kinds of results in the long tails of paths are not possible in small and sparse networks. Alg-cc curves in Fig. 13 with \( L_{\text{max}} = 8 \) and \( L_{\text{max}} = 11 \) reveal also network structures that are not visible in the Sim-sc curve. In the earlier work (P I) I investigated the effects of peripheral nodes by comparing the present model that allows loops with a model containing only self-avoiding paths.
Figure 12. A Facebook social media network of 4039 nodes. Nodes represent users in the network and link connections between them. Node sizes in the graph describe the centrality values of nodes.

Figure 13. The values of out-centrality and in-centrality measures of the Facebook social media network of Fig. 12 with link weights $w_l = 0.05$ for Alg-cc and Sim-sc. Alg-cc results are displayed for the maximum path lengths $L_{max} = 8, L_{max} = 11$ and $L_{max} = 13$. Alg-cc show complex contagion results and Sim-sc show simple contagion results. The order of source nodes (horizontal axis) is determined in the ascending order of Sim-sc out-centrality or in-centrality values.
5.3 Ego-centric Networks

The model of influence spreading can also be used for analysing human behaviour and interaction with others in social networks. In this section, this model and corresponding influence centrality measures are applied to real data of mobile phone call detail records. To analyse 48,000 egos in their ego-centric networks the out-centrality, in-centrality and betweenness measures are used. The influence spreading in the ego-centric networks is analysed under different scenarios of link strengths between individuals reflecting the network structure either being interaction or connectivity oriented. The model reveals the characteristics of social behaviour that are not obvious from the data analysis of raw empirical data or the results of standard centrality measures. A transition is discovered in behaviour from young to older age groups for both genders and in both normalised out-centrality and in-centrality as well as in betweenness centrality results. The following functional form is used to describe the strengths of social ties for the link strengths from node $i$ to node $j$:

$$w_{ij} = 1 - (1 - \alpha)^{n_{ij}},$$

(25)

where $n_{ij} \in N$ is the number of communication events from node $i$ to node $j$, and $0 \leq \alpha \leq 1$ is a tuning parameter. The parameter $\alpha$ determines two different classes (or facets) of the influence spreading processes through the network. The first class, called a “connectivity” facet, is expressed in the range $\alpha \approx 1$, where the link weight factor $w_{ij}$ is close to 1. In this case, the definitions for centrality in Eqs. 19–20 and betweenness in Eq. 21 behave like the classical definitions, such that the connectivity of the network plays a central role in the influence spreading process. On the other hand, in the range $\alpha \approx 0$, the number of pairwise interactions $n_{ij}$ plays an important role, as the value of $w_{ij}$ depends strongly on $n_{ij}$. This is called an “interaction” facet.

Next, I demonstrate how new metrics can be defined by combining two different measures, the degree and the influence centrality measures of Eqs. 19–20, to highlight specific characteristics of influence spreading in the ego-centric networks. The in-degree and out-degree centrality measures depict the general trend of degree centrality, but there are differences between the influence in-centrality and influence out-centrality results when the parameter $\alpha$ of the link weight model is varied. The influence in-centrality and out-centrality are scaled with the in-degree and out-degree, respectively, to get the results for the normalised centrality measures, as presented in Fig. 14. These normalised quantities show very different features depending on the influence facet, i.e. for different values of $\alpha$.

The normalised influence centrality represents a novel measure of the influence that individuals experience from and exert on their social network. The measure unveils interesting and different dynamics depending on the influence facet applied. One observation is that in the interaction facet normalised in-centrality is increasing and normalised out-centrality is decreasing for older than 40-year-old groups for both genders. These trends turned out to be much weaker or non-existing in the connectivity facet ($\alpha \approx 1$). The interpretation is that for the older age groups there is a shift in the interaction-based social
behaviour but not in the connection-based behaviour. The youngest age groups under 30-year-old are both more influential in their social network and more influenced by their social network when compared to the older cohorts. This behaviour holds in the interaction facet ($\alpha \approx 0$) and the connectivity facet ($\alpha \approx 1$).

In the interaction facet, a maximum around the 30-year-old age group is shown particularly in the normalised out-centrality. In the interaction facet, the cohorts in middle adulthood tend to be more relevant for the influence spreading process than other age cohorts, such that females have a larger influential role than males throughout all ages. The youngest age cohorts have a somewhat smaller influential role in the spreading process, which could be understood from the fact that these cohorts have recently entered the economic/labour part of the society and they are still in the process of generating a relevant role in it. The younger age groups have almost as high normalised in-centrality as 30-year-old. In the interaction facet, female egos have both larger normalised in-centrality and out-centrality values.

In the connectivity facet young male cohorts under 40-year-old have larger normalised in-centrality values than females of the same age. The normalised out-centrality in the connectivity facet shows a clear drop around the 20-year-old age group and for older than 30-year-old egos the normalised out-centrality is almost a constant. The normalised out-centrality in the connectivity facet behaves similarly for both genders.

In addition, the betweenness centrality has been determined as a function of the ego’s age and for both genders for different values of the parameter $\alpha$. In Fig. 15 the results for $\alpha = 0.01$ and $\alpha = 0.5$ are shown. In the interaction facet, females have considerably larger betweenness centrality than males, whilst in the connectivity facet, there are no significant gender differences. It is also seen that the betweenness centrality for $\alpha = 0.01$ again shows a prominent peak at about 30-year-old age cohort for both genders but females with higher values than males throughout all ages. For $\alpha = 0.5$ the peak gets less prominent and tends to disappear, and the difference between females and males dissolves and gets slightly reversed for 40-years and older age cohorts. The betweenness centrality grows gradually for the subsequent age cohorts until the age of 40-year, after which the relevance of each group is practically the same, though a bit larger for males than females. This observation, in terms of the connectivity facet of the influence spreading process, could show that the role of individuals in their local networks as bridges for influence spreading, is established in the first stages of adulthood, after which a well-established relevance is preserved for the subsequent ages.
Figure 14. The normalised influence in-centrality (on the left) and out-centrality (on the right) values for males (blue) and females (red) as a function of the ego’s age and the link weight model in Eq. 25 with parameter $\alpha = 0.01$ (top) and $\alpha = 0.5$ (bottom). The curves for $\alpha = 0.01$ and $\alpha = 0.5$ represent interaction and connectivity facets respectively.

Figure 15. Egos’ betweenness centrality values (Eq. 21) as a function of ego’s age for males (blue) and females (red). The averages are calculated for two different values of the link weight parameter $\alpha = 0.01$ and $\alpha = 0.5$ shown in the left and right panels. The curves for $\alpha = 0.01$ and $\alpha = 0.5$ represent interaction and connectivity facets respectively.

5.4 Brain Network

Next, I will discuss supplementary or alternative community detection methods to hierarchical clustering methods that have been previously used in the literature to investigate brain networks [47]. I will present modelling methods and possible new properties of complex network subsystem cooperation and organisation.
The centrality and betweenness measures are useful for discovering and analysing hubs in network structures. In our influence spreading model, two centrality measures can be defined for output and input interactions in Eqs. 19–20 while there is only one version of the betweenness measure in Eq. 21. These measures characterize different roles of hubs in the brain network. The out-centrality measure is designed for describing the power of a node’s influence on other nodes in the network. The in-centrality measure describes the influence of other nodes on a node in the network. In Fig. 16, in-centrality measures are indicated with blue and out-centrality measures with orange bars. The numerical values of in-centrality and out-centrality are shown for 122 regions (nodes) of the brain network.

Names of the regions are indicated as labels to enable comparison with the original research article [47]. Two digits in the labels indicate the two cerebral hemispheres and the four modules detected by the Multiresolution Consensus Clustering Method (MRCC) algorithm [47]. One observation is that the values of in-centrality and out-centrality are unbalanced in many cases: in-centrality values are high for {11 NDP, 21 NDP, 13 PERI, 13 PL, 13 BLAa, 13 ORBm, 13 ORBv, ...} and out-centrality values are high for {11 ACAd, 13 CP, 13 ACB, ...}. In-centrality and out-centrality measures of the influence spreading model provide more information than the standard node degree results presented in [47].

The betweenness measure is a compromise between the two centrality measures. The betweenness measures the role of a node as a mediator of influence, and the relative impact on the influence spreading process if a node is removed from the network. Figure 17 shows the values of the betweenness measure as defined in Eq. 21. The centrality measures and the betweenness measure are correlated but interesting differences can be observed in the details.

Next, I will briefly discuss some representative results of the community detection method. The algorithm optimising the quality function in Eq. 24 discovers seven different divisions of the entire brain network [47]. The regions (nodes) are connected and cooperate functionally in some way in the brain. These kinds of interactions between subsystems can be important in analysing subsystem organisation. Different views can exist simultaneously and they can be useful in understanding different aspects of complex systems such as brain networks. This kind of analysis can also reveal subsystem cooperation between the two hemispheres.

For comparison, I calculate whether two different network models, our influence spreading model (See Eq. 17) and the standard network connectivity model (See Eq. 11), lead to different subsystem structures. Subsystems are sets of regions (nodes) in the brain network that optimise a quality function. These results are close to each other but differences in details exist particularly in nodes near boundaries (predicted by the Gephi software) of the left and right partitions. The conclusion is that the main results are similar but interaction characteristics can have an impact on detailed results.
Figure 16. The values of the in-centrality and out-centrality measures for the 122 first regions (nodes) of the brain network.

Figure 17. The values of the betweenness measures for the 122 first regions (nodes) of the brain network.

Figure 18 shows the main division into two hemispheres. In addition, the second-order division is indicated with colours. In Fig. 18 individual regions cooperating across the boundaries of the main division are collected and highlighted with large font size numbers. The high-level layout of Fig. 18 generated by the Gephi software coincides very well with the community detection algorithm in Eq. 24 and also with the results in [47].

Examples presented in this section illustrate how cooperation among subsystems across module and hemisphere boundaries can be studied. Assuming that particular types of network processes describe cooperation among subsystems, such as connectivity or a spreading model, can lead to different results and conclusions. If mechanisms and network processes are not completely understood, alternative models can provide new insights. Different models are useful in confirming or rejecting alternative presumptions about systems' behaviours.
5.5 Epidemic Spreading

Here, I propose a mathematical model for describing epidemic spreading in structured organisations. Our motivation arises from the ongoing Covid-19 pandemic but our goal has been more general in developing a flexible modelling methodology for different spreading dynamics.

The Covid-19 virus spreads dominantly via respiratory droplets and aerosols from an infected person but it can spread also by physical contact. Covid-19 is an example of a virus that spreads via social contacts and people gathering together. People with strong social ties more frequently get together, usually indoors, and consequently have a higher probability of spreading and contracting the virus. The model is based on using bi-directional links with weights that describe probabilities of spreading infection between two individuals during the infectious period of a disease. The additional effect of frequent gathering among close friends is taken into account by the model’s repeated and circular interactions in the social network structure.

One aspect integrated into the ways of epidemic spreading is that breakthrough infections via people recovered from infection and vaccinated individuals can also transmit infection. What makes the problem even more complicated is that differentiating the phenomena in empirical studies may be difficult and it is possible that spreading through respiratory droplets and aerosols is partly miscounted as breakthrough infection or vice versa. The magnitude of the breakthrough effect is notable but varying estimates of the percentage have been published in the literature [48]. Social interactions, spreading through the air
and breakthrough infection work as the guidelines for developing a model of epidemic spreading.

Modelling an epidemic spreading in a population or an organisation is important in planning preventive measures. I will focus on investigating the effects of limiting contacts between different organisation structures and shortening chains of infection together with general risk mitigation actions. I calculate the effects of combined measures that enable the policy-makers to select the most effective and achievable set of concurrent actions. Out-centrality, in-centrality and betweenness measures are used in analysing different aspects of epidemic spreading. Also, examples of calculating community immunity are presented, in which strategies based on out-centrality and betweenness measures are prioritised. I conclude that preventive measures can have a significant effect on achieving community immunity.

As an example, I will use the modular network in Fig 17 to describe a real-world organisation structure. Four different types of interactions between individuals are studied: leadership or management, contacts among individuals from the same department, close interactions with roommates and other kinds of interactions between departments. I have analysed the impact of interventions weakening different types of contacts on several epidemiological indicators such as the reproduction number [49], the percentage of infected individuals or the prevalence needed to reach herd immunity [34]. The structural spreading model (P I; P II) is used to estimate the effects of governing contagions inside the organisation and the important role of bridge nodes in the diffusion of epidemics across different departments in the organisation. I have estimated the effect of weakening different contacts in the organisation compared to the case where no such measures are taken, and also compared different measures and their combinations. This kind of information is valuable when selecting the most effective measures for confining the spread of an epidemic.

Preventive measures in public and private organisations include limiting contact between members of the organisation and with the outside society. I have categorised contacts into four main classes and studied the effects of limiting contacts between organisation structures. The four classes of relationships are used to illustrate the proposed model. The model allows defining any number of different groups or even individual parameters for all the nodes and links of the organisation structure. Quantitative results of the model can be used for planning the most effective combination of preventive measures.

<table>
<thead>
<tr>
<th>Preventive measures</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>q</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>a</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>r</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>z</td>
<td>0.01</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 19. The network view of connections in the organisation structure. Peer (or roommate) relations are shown with the black colour and other close relations with the blue colour. Nodes 13—14, 54—55, 95—96 and 136—137 are bridge nodes between administrative staff and managers inside the four departments. Dashed lines show the borders of the administrative staff and the four departments.

In epidemiology, the basic reproduction number $R_0$ [49] of an infection is the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection. The definition assumes that no other individuals are infected or immunised naturally or through vaccination. For the first variant (Alpha) of the Covid-19 virus, the basic reproduction number in the population was estimated to be approximately $R_0 = 2.2$ [33, 50]. Here, the basic reproduction number in the society is denoted by $R_0$ and the basic reproduction number in the present example organisation is denoted by $r_0$. The relation between the model parameter value of $W_I$ and $r_0$ is linear as can be seen in Fig. 20 for the case of no preventive measures (−) and preventive measures ‘z’, ‘r’, ‘a’, ‘z, r’ and ‘a, z, r’. One can see from Fig. 20 that in the organisation structure of Fig. 19 with the parameter values of Section 5.5 the maximum value of $r_0 = 13.3$ with no preventive measures, and $r_0 = 11.9, 10.2, 9.2, 8.9$ and 4.8 for the five preventive measures respectively.
When modelling the effects of different preventive measures against infectious disease in an organisation, the number of incoming spreaders from the outside society is one approach to studying epidemic dynamics and control. I assume that 1–3 members of the organisation arrive carrying the infection or 1–3 visitors import infection to the organisation. Figure 21 (top) shows spreading probabilities without preventive measures as a function of the number of spreaders for parameter values $W_l = 0.22, 0.18, ..., 0.06$ (from top to bottom). Figure 21 (bottom) shows the effects of limiting contact between departments and also between departments and administrative staff.

Curves in Fig. 22 show the results for no preventive measures (top) and with the preventive measure for cross-department contacts (bottom) for different path lengths. Path length $L = \infty$ indicates that the results are for long path lengths, where increasing path lengths in the computations no longer increases spreading probabilities, and an equilibrium state has been reached. Coloured curves show corresponding results for path lengths $L = 3, 5, 7$ and 9. The results show the impressive effects of restricting infection chains in both cases. In real situations, with the help of testing and tracing, it is possible to shorten the chains of infection.
Figure 21. Percentage of infected nodes as a function of the number of initial spreaders (#) in the organisation of Fig. 19 for no preventive measure (−) and contacts limited between departments (z). The parameter values of $W_l$ for the seven curves from top to bottom are $W_l = 0.22, 0.18, 0.14, 0.12, 0.10, 0.08$ and $0.06$.

Different metrics can be used to analyse spreading power, probability of contracting infection and effectiveness of transmitting infection between members of the organisation. These characteristics can be measured by out-centrality, in-centrality and betweenness measures of Eqs. 19–21. In Fig. 23, I show some results of these three different centrality measures in the case of the organisation structure of Fig. 19. A node’s out-centrality measures how effective the node is in spreading infection in the network. In-centrality measures the probability of a node contracting the infection from the network. The betweenness metric measures the effectiveness of a node to transmit infection between other nodes of the network.

Combining results from out-centrality, in-centrality and betweenness metrics can provide insight into the analysis of epidemic dynamics in a structured organisation. Preventive measures and their concurrent use can be compared and analysed in detail from different perspectives.
Community immunity, or herd immunity, is achieved when a sufficiently high percentage of the population is immunised so that the basic reproduction number is less or equal to one \( r_0 \leq 1 \). Immunisation is attained by vaccination or by recovering from an infectious disease that gives immunity. Fig. 24 shows that community immunisation is achieved when about 48.1% of the members of the organisation in Fig. 19 are immune (solid curve top figure). Only about 43% is needed for community immunity with the preventive measure for cross-department interactions (solid curve bottom figure). The level of \( r_0 = 1 \) is indicated by the solid green line in the figures. Linear relations in Fig. 20 can be used to estimate corresponding percentage values for preventive measures ‘\( r' \), ‘\( a' \), ‘\( z \), \( r' \) and ‘\( a, z, r' \) as 37%, 33%, 32% and 17% respectively.

Figure 22. Percentage of infected nodes in the organisation in Fig. 19 for different path lengths (top) and the same path lengths with the preventive measure of limiting cross-department contacts (bottom).
Out-centrality and betweenness metrics can be used to prioritise who should be vaccinated or isolated in the network. A lower percentage is needed for community immunity: about 43 % and 36—39 % for the two cases of no preventive measures and the preventive measure of limiting cross-department contacts, respectively. In Fig. 24 dashed curves show the effects of prioritising nodes having...
the highest out-centrality (blue) and lowest out-centrality (red). The dotted curves show the corresponding effects when using the betweenness metric. Prioritising individuals with low values of out-centrality and betweenness metrics gives the worst scenarios. These results are for informative purposes only. The out-centrality performs better than betweenness but there is not much difference between these two results in the top figure with no preventive measures. Particularly, with the preventive measure of limiting cross-department contacts, the betweenness metric does not perform as well as out-centrality for prioritising the order of vaccination or for selecting individuals into isolation.

In a special case of the model, when the vaccinated and recovered individuals are immunised, the results have been obtained using a simple contagion model. The model is more suitable for modelling diseases where the disease spreads from person to person and immunised individuals do not transmit the disease. In the model, this means that the maximal number of visits on a node is restricted to one and loops are not allowed in the chain of transmission.

When comparing the results of the complex contagion model and the simple contagion model, I find that preventive measures have different effects, and in some cases, the prioritising of preventive measures is changing. The optimal preventive measures can vary according to different virus types or even virus variants. These results are for our example organisation structure in Fig. 19 and specific parameter values (Section 5.5) but similar effects can be found in other cases. The idea of using the model is first to construct the organisation structure and then evaluate the spreading probabilities for different groups or individuals for the actual virus variant [48, 51, 52] in that environment.

In summary, I have presented a complex contagion spreading model and proposed new metrics for describing epidemic spreading in structured organisations. The model enables detail-level modelling of spreading probabilities between nodes of the social network structure.

We describe different aspects of spreading mechanisms: direct person-to-person contact and spreading through the air in close contact. In addition, the breakthrough infection can be a method of spreading via vaccinated or recovered individuals. The spreading mechanisms of Covid-19, for example, involve all of these mechanisms but their exact proportion is not yet known. In our current model, repeated and circular spreading dynamics on the contact network structure are used to model the relatively higher spreading probabilities of close contact environments.
Community immunity without preventive measures (—) is achieved when about 48.1% of the members of the organisation in Fig. 19 are immune and about 43% with the preventive measure of limiting cross-department contacts ($'z'$). Out-centrality (dashed curves) and betweenness (dotted curves) show lower and upper bounds when these metrics are used for prioritising nodes in the network in Fig. 19.

**Figure 24.** Community immunity without preventive measures (—) is achieved when about 48.1% of the members of the organisation in Fig. 19 are immune and about 43% with the preventive measure of limiting cross-department contacts ($'z'$). Out-centrality (dashed curves) and betweenness (dotted curves) show lower and upper bounds when these metrics are used for prioritising nodes in the network in Fig. 19.
Figure 25. Percentage of infected individuals for one initial spreader (index case) in the organisation in Fig. 19 for no preventive measures (−) and preventive measures ‘z’, ‘r’, ‘a’, ‘z&r’ and ‘z&r&a’ in the complex contagion (CC) model (top) and the simple contagion (SC) model (bottom).
6. Scientific Results

Primary research question (See Section 2.1): “How to develop an influence spreading model that describes the detailed network structure of a network that considers all paths between nodes in the network?”

In this thesis, a novel influence spreading model is proposed to describe spreading processes on a detailed network structure. The network structure is defined as nodes and directed links between nodes. Node and link weights are used to model the probabilities of spreading over a node and through a link respectively. The modelling framework consists of constructing the network structure and describing dynamic spreading processes on the structure. The spreading process is initiated from a node or nodes in the network and influence propagates via links between nodes. In a small network structure, it is possible to consider all paths in the network but in larger networks computing time is limited by using the model parameter $L_{\text{max}}$ for the maximum path length. Here two main models for complex contagion and simple contagion processes are proposed. The complex contagion model allows loops and breakthrough influence via nodes. The simple contagion model does not allow loops or breakthrough influence. These models have applications in modelling social networks and epidemic spreading.

Research sub-question RQ1: “How can we define generalised centrality and betweenness measures in the model?”

Novel metrics for closeness and betweenness metrics are proposed based on the spreading probabilities calculated from the spreading model. The spreading probabilities are expressed with an $N \times N$ influence spreading matrix for a network of $N$ nodes. Closeness centrality measures can be defined for both out- and in- directions for the nodes in the network.

Research sub-question RQ2: “How can we detect communities and sub-communities in the model?”

Here a novel community detection method based on the influence spreading matrix is proposed. The method considers all paths in the network within the maximum path length. This differentiates my model from most of the commu-
Community detection methods proposed in the literature. The method detects overlapping and hierarchical structures. The node and link weights can be used to describe weak and strong interactions or different levels of granularity of results.

Research sub-question RQ3: “How can we define quality measures (objective functions) for the detected communities in the model?”

The present community detection method uses a simple form of an objective function based on the influence spreading matrix. The method searches the local maxima of the quality function detecting different splittings of the network. The form of the objective function is simpler than the standard modularity function often used in the literature. The objective function that I have used in the algorithm for detecting communities and sub-communities can be used as a quality function to compare the cohesion between any sets of nodes in the network structure. In addition, I have proposed alternative quality functions for measuring the probability of formation and robustness of composition for different community structures.

Research sub-question RQ4: “How can we incorporate different temporal distributions into the model?”

Different temporal distributions can be implemented in the spreading model. Temporal spreading has been demonstrated with two different distributions, the Poisson distribution and a probabilistic distribution describing temporal delays in message forwarding events.

Research sub-question RQ5: “How to apply the model for describing epidemic spreading?”

The epidemic spreading in structured organisations is investigated using two versions of the model, one with complex contagion and the other with simple contagion. The complex contagion model allows loops and breakthrough infection via nodes. The simple contagion model does not allow loops or breakthrough infection. In addition, a model for partial breakthrough infection is proposed where a model parameter is used to control the fraction of the breakthrough effect. I assume that the complex contagion model can describe the spread of virus infections where the main method of spreading is through the air and breakthrough infection has a large effect.

Table 5 summarises publications and the most important formulas where my research sub-questions have been discussed and answered. The first research sub-question RQ1 regarding the influence spreading model has been covered in all publications PI–PVII. The research sub-questions RQ2 and RQ3 have been considered in publications PI, PII, PIII and PVI and the research sub-question RQ4 have been considered in publications PI, PII and PIII. Finally, the research sub-question RQ5 has been covered in publication PVII.
Table 5. Research sub-questions RQ1—RQ5 have been answered in Publications P I—P VII. In addition, I have indicated the most important formulas in the publications related to the research sub-questions.

<table>
<thead>
<tr>
<th>RQ/Publ</th>
<th>P I</th>
<th>P II</th>
<th>P III</th>
<th>P IV</th>
<th>P V</th>
<th>P VI</th>
<th>P VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Eqs. 9—14</td>
<td>Eqs. 5—7</td>
<td>Eqs. 5—9</td>
<td>Eqs. 3—7</td>
<td>Eqs. 3—7</td>
<td>Eqs. 1—7</td>
<td>Eqs. 1—7</td>
</tr>
<tr>
<td>RQ2</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eq. 15</td>
<td>Eq. 8</td>
<td></td>
<td></td>
<td>Eq. 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RQ3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eq. 15</td>
<td>Eq. 8—9</td>
<td>Eq. 10</td>
<td>Eq. 10</td>
<td>Eq. 10</td>
<td>Eq. 10</td>
<td></td>
</tr>
<tr>
<td>RQ4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eq. 8</td>
<td>Eq. 10, Code in P II</td>
<td>Eq. 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RQ5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eqs. 1—7</td>
</tr>
</tbody>
</table>


7. Future Research

There are a number of areas of future research related to this study. I list some of them in the following:

1. Selecting optimal team members from a social network
2. Community detection based on locally dense edges
3. Signed networks
4. Dynamic network structure
5. Different temporal spreading distributions
6. Different network models
7. Comparison of community detection results and centrality measure values with different models in the literature

The community detection method of Section 4.4. can be modified to select optimal team members from a social network. An appropriate objective function for this purpose can be developed if there are specific tasks, roles, requirements or priorities in the team. Alternatively, the optimised quality function could be defined by using pairwise metric values of centrality and betweenness measures in Eqs. 19–21. One aspect in selecting the team members is whether they have or need support for their tasks from other individuals or groups outside the team.

Related to the problem of selecting optimal team members is a community detection method that is closer to the classical optimisation of modularity where communities are detected based on locally dense weighted edges in the network structure. This method answers a different question and applies to different applications. The method proposed in Section 4.4 assumes that a network structure is split into two halves where both fractions compose a community with social ties between their members. This is the case in many applications. Sometimes coherent groups can build up spontaneously around an idea or ideology inside a larger network structure, where the outsiders do not have a similar connecting or unifying factor. Note that the current model in Section 4.4. reveals sub-communities because typically many local maxima are found with the method (P I; P II; P III; P VI). Experimenting with a modified method, where the splitting is conducted into more than two factions, yielded results that were close to the results of the current method. Publishing these results after further analysis can be considered a future work. Here again, if the community formation process could be known better, the model could be adjusted accordingly.
In a social network analysis and specifically in structural balance theory [53], a positive or a negative tie can be established between two nodes in a network. People can be friends but also foes. This results in a signed network. The community detection algorithm of Section 4.4. could be modified to describe signed networks simply by searching local minima instead of local maxima in Eq. 24.

Many applications have dynamic network structures such that nodes or links can disappear or emerge during the spreading process. Another application has time-dependent node or link weights or the weights change according to a predefined rule depending on the spreading process development or specific characteristics in the network structure.

The dynamic network structures can be studied more extensively by investigating different temporal spreading distributions. In this thesis, only two probability distributions are demonstrated (P II). Also, empirical temporal spreading distributions can be incorporated into the model and even individual distribution functions could be used for individual nodes or connections in the network.

In this thesis only two different network models have been discussed: the spreading model (P II), with its complex contagion and simple contagion variants (P VII), proposed in this study, and the classical network connectivity model by [28]. Many other possibilities and combinations can be useful in describing complex interactions, particularly in social networks or in epidemiological models.

The results of the models in this thesis should be compared with the results of other models and studies in the literature. I have proposed novel closeness centrality measures and a betweenness measure that belong to the same modelling framework as the influence spreading model. These metrics can be compared with corresponding quantities in the literature. Maybe hundred of different community detection methods have been proposed in conference and journal articles. The results of the community detection method could be compared with the most popular methods in the literature. Comparing the results of different metrics or algorithms includes also discussing processes and applications that they are intended to describe. For example, as explained in Section 1.3, Burt’s structural holes [1] and Granovetter’s weak ties [2] describe different social interactions than Centola’s complex contagion processes [3].
8. Conclusion

In this thesis, I have proposed a novel influence spreading model and metrics based on the spreading model to be used in different applications from social networks to epidemic spreading. I have studied how to better characterise the important nodes or communities on complex networks for different transmission processes. The basic idea is to decompose the analysis into two steps. The first step is to calculate the strength of the influence from a node to another node from the given transmission process for every directed node pair which gives the so-called influence matrix. Then the centrality measures and community structures are calculated from the influence matrix. The main model describes processes on network structures that can be characterised as complex contagion processes. A special case that does not allow loops or breakthrough influence via nodes is called a simple contagion model.

Furthermore, I have identified several potential areas for future research, for example, there is a need to develop the model’s features for different types of interactions. For example, breakthrough influence in social networks, and corresponding breakthrough infection in epidemic spreading, is important to describe some processes on networks more realistically. However, the issue is more general than just how to model the network processes: quantifying social and epidemiological interactions and defining concepts are parallel lines of research.
13. B. Min and M. San Miguel, "Competing contagion processes: Complex contagion triggered by simple contagion,” *Scientific Reports*, vol. 8, 10422, 2018. [https://doi.org/10.1038/s41598-018-28615-3]


Network Science is a growing field of research, and it has gained a lot of interest in many areas of applications. One definition of network science is the study of network representations of physical, biological, and social phenomena aiming at predictive models for them.

In this study, I focus on investigating influence spreading and present a model that can be used in various applications to study different spreading processes in network structures. Applications cover describing influence spreading and community formation in social networks, information transfer in communication networks, transportation in infrastructure networks, and epidemic spreading in populations.

Detailed level descriptions of the network and its elements are needed in calculating various centrality measures for network elements and objective functions for community formation. The goal is to develop a common framework with consistent mathematical concepts, metrics and methods. All the models and measures proposed in this study are based on probabilistic quantities and thus have physical interpretations.