
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Kuikka, Vesa; Kaski, Kimmo K.

Detailed-level modelling of influence spreading on complex networks

Published in:
Scientific Reports

DOI:
[10.1038/s41598-024-79182-9](https://doi.org/10.1038/s41598-024-79182-9)

Published: 01/12/2024

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY-NC-ND

Please cite the original version:
Kuikka, V., & Kaski, K. K. (2024). Detailed-level modelling of influence spreading on complex networks. *Scientific Reports*, 14(1), 1-21. Article 28069. <https://doi.org/10.1038/s41598-024-79182-9>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



OPEN Detailed-level modelling of influence spreading on complex networks

Vesa Kuikka¹✉ & Kimmo K. Kaski^{1,2}

The progress in high-performance computing makes it increasingly possible to build detailed models to investigate spreading processes on complex networks. However, current studies have been lacking detailed computational methods to describe spreading processes in large complex networks. To fill this gap we present a new modelling approach for analysing influence spreading via individual nodes and links on various network structures. The proposed influence-spreading model uses a probability matrix to capture the spreading probability from one node to another in the network. This approach enables analysing network characteristics in a number of applications and spreading processes using metrics that are consistent with the quantities used to model the network structures. In addition, this study combines sub-models and offers a comprehensive look at different applications and metrics previously discussed in cases of social networks, community detection, and epidemic spreading. Here, we also note that the centrality measures based on the probability matrix are used to identify the most significant nodes in the network. Furthermore, the model can be expanded to include additional properties, such as introducing individual breakthrough probabilities for the nodes and specific temporal distributions for the links.

Keywords Network spreading model, Social network, Complex network, Computational social science, Community detection, Centrality measure

Network science is a research approach that by means of data analysis and computational modelling investigates the structural and dynamical properties of complex systems by considering them as networks of distinct elements or actors and connections between them, i.e. of nodes and links^{1–3}. This approach can be applied to any system with dependencies between its components, like social networks, telecommunication, transportation, power grids, semantic networks, and biological networks. In addition to networks' structural and dynamical properties network science approach offers us ways to study the resilience of a network, which has become a crucial issue for societal infrastructure and communication networks². In case of social networks, one focuses on investigating social relations and their temporal changes between people, linked by friendships, family ties, professions, hobbies etc. As a result of the growing role of social media services³ in our lives and social networks, understanding of and insight into how knowledge, behaviour, cultural resources, opinions, and generally speaking influence spreads is increasingly important.

For deeper understanding of influence spreading in human social networks, computational modelling serves as a versatile research tool, due to the fact that it can cover both continuous and nominal traits, including opinions. Here the opinions can be considered as a broad concept that encompasses personal beliefs, behaviour, and attitudes⁴. In general dynamical phenomena on networks are often modelled using the random walk approach, but their analytical solutions can be difficult or infeasible for complex topological structures and temporal networks⁵. Here we propose an influence-spreading model that can be expressed analytically, though writing down the formulas for complex network structures would be impractical. Instead in this study, we will take approaches related to Markovian methods, matrix algebraic network models and mean-field approaches and introduce a novel probabilistic method to calculate the spread of the influence and the derived results. Our model for influence-spreading describes a stochastic process in which the future state of the network depends only on the present state, known as the Markov process. However, modelling detailed-level spreading processes on structured networks is not practical with standard Markov chain methods whereas mean-field methods approximate the original interactions with an average or effective interaction.

When it comes to model influence spreading on social networks, there are a few key concepts to keep in mind. These include the network structure, link and node weights, paths of influence, and spreading rules.

¹Department of Computer Science, Aalto University School of Science, P.O. Box 15500, 00076 Aalto, Finland. ²The Alan Turing Institute, 96 Euston Rd, Kings Cross, London NW1 2DB, UK. ✉email: vesa.kuikka@aalto.fi

The network structure is defined by the directed links between adjacent nodes, with each link having a weight value assigned to it as a spreading probability. Here the paths of influence refer to lists of nodes in the network structure, the influence spreading starting from a particular node and travelling through various paths to an end node in the network. Our model is classified as a non-conserved spreading model where spreading from a node continues to all adjacent nodes according to the link weights. For computation efficiency or phenomenological reasons, the path length can be limited to a maximum path length⁶. It is important to note that small link weights, representing a low probability of relaying influence, can cause a rapid decrease in spreading probabilities on individual paths. However, combining the effects on end nodes through different paths has the opposite effect, which could be amplified if circular paths are allowed during the spreading process^{6–8}.

There are two main classes of spreading processes, namely complex contagion and simple contagion activation of nodes. In the standard definition of complex contagion, an individual requires multiple exposures to innovation before adopting the behaviour change, while in simple contagion innovation may spread in one contact with an influenced neighbour⁹. These definitions of simple and complex contagion may be too restrictive, as the standard definition of complex contagion⁹ considers only the neighbouring nodes and overlooks the network spreading processes. Also, recurrent events are not considered in the standard definitions of simple and complex contagion. In this study, we will generalise the standard definitions such that for complex contagion we allow recurrent events between node pairs and circular events or loops on paths, while for simple contagion we only allow self-avoiding paths where one node cannot appear more than once on a path. This generalisation of standard definition allows different network spreading processes to occur.

In our previous studies, we presented an algorithm to model influence spreading in a network by introducing individual node and link probabilities, expressed in terms of an influence spreading matrix with spreading probabilities from one network node to another. Here, all quantities are expressed as probabilities, which means that the derived results, such as centrality measures and quality functions for community detection, are commensurable. This approach serves as a framework for comprehensive network analysis made possible by the detailed-level modelling of network structures and influence-spreading processes. It can also be used to showcase novel features of the model, such as temporal distributions and short-term effects based on path lengths.

The main focus of this study is to use the framework of the probabilistic influence-spreading model to analyse influence spreading in networks and simultaneously develop community detection, centrality and betweenness measures of the influence network. Hence the influence-spreading model enables us to analyse both the spreading process in the network and its structural properties consistently. In addition, as the probabilities generated by this model are used to express the network centrality measures and the objective function for community detection, these quantities can be analysed together to produce new results. To emphasise the importance of a single general modelling framework instead of just using point solutions we will make a case study of our influence-spreading model. We believe that network analysis needs more generic methodologies for the wider application use, instead of or complementary to the current situation in which, for instance, network metrics and community or modularity detection methods have been developed independently of each other^{1,10–13}.

Our overarching motivation for this study is to demonstrate the modelling approach and its benefits gained when link weights of the influence-spreading model and derived quantities are expressed with the same units of probabilities. Multiple methods working together often achieve better comprehensibility and interpretability of the results than a single method alone. Our methodology allows for simultaneous analysis of network structures from different perspectives, providing compatible methods and metrics for structural network analyses. Many of the methods and metrics commonly used for analysing social networks have been developed independently, and they are not directly comparable in numerical value, which can limit their simultaneous use in applications. This is why, our aim is to promote general modelling methods where network processes, network metrics and community quality functions are considered on the same basis to enable more consistent and deeper analysis of the results. This approach can provide new insights and results across different applications of network science, specifically in the modelling of spreading processes.

We have chosen to use the Les Misérables network in Fig. 2a to showcase the applications of our model. This network is a relevant example because a recent research article¹⁴ used an information-theoretic method to identify groups of nodes that often appear together in the same community, and we can compare those results with the ones of our model. The research method¹⁴ is significant because it employs general information theory rather than relying on a specific community detection method from the existing literature. In addition, the Les Misérables network structure is suitable for demonstrating centrality measures.

While the proposed model has great extensibility due to its detailed-level design, there are certain limitations. Firstly, not all potential model features have been implemented in the current version. Secondly, despite the scalable Algorithm (see [Supplementary information](#) online), the detailed level computation increases the time complexity of the model, which is high for some contemporary applications such as big data and large networks. The third limitation pertains to the theoretical constraints of the analytic algorithm of the complex contagion. The current version of the computer program is designed for the non-mutually exclusive events of basic probability theory and full breakthrough effects. Our implementation for breakthrough probabilities less than one has been executed as a simulation computer program¹⁵. It remains a research question as to how more efficient methods, whether analytical or simulation-based, can be designed for state-dependent processes. Lastly, the algorithm utilising the theory of non-mutually exclusive events in probability theory may not be extendable for all processes on network structures. The critical requirement is that the method is based on the assumption that alternative propagation paths can be considered in any order in the algorithm. If this condition does not hold, it must be taken into account in the algorithm or an approximate solution is necessary. We will discuss model limitations in a particular application in the context of epidemic spreading.

In the next Sect. 2 we explain the model design principles of probabilistic influence-spreading. In Sect. 3 we discuss the research work related to our study, namely Markovian models, higher-order contagion models, standard network centrality measures, and community detection methods. Next in Sect. 4 we present the algorithms with formulas of our influence-spreading model in detail. Then in Sect. 5 we provide numerical examples to demonstrate the usefulness of our model in its applications, first in the present use case of a small Les Misérables network structure (see Fig. 2a) then followed by cases presented in our previous studies analysing spreading processes in networks of different sizes including influence spreading and opinion formation, epidemic spreading, information spreading, and network resilience^{6–8,15,16}. In Sect. 6 we summarise our study and discuss its limitations.

Design of probabilistic influence-spreading model

In our earlier study, we demonstrated that the Algorithm (see [Supplementary information](#) online) based on the probability theory approach to influence spreading ensures accurate calculation of spreading through alternative paths in network structures of different sizes and complexity. It has also been found that detailed-level modelling approach is essential for modelling epidemic spreading and countermeasures at the organisational level¹⁶ and in the case of small-scale structures or overlapping communities¹⁷. The design principle for modelling influence spreading on a detailed-level network structure enables us to meet two fundamental requirements of the network analysis methods, namely describing network flow processes and defining higher quality unbiased centrality measures. The need for process-driven network analysis is discussed in¹⁸, and the need for high-quality centrality measures in¹⁰. Our approach addresses both these needs by modelling two basic contagion models, simple and complex contagion, and proposing novel forms of unbiased centrality measures. A consequence of our detailed-level modelling approach is that the model has great potential for extension to consider different processes and possibility to introduce application-related factors. Breakthrough effects and temporal spreading distributions are examples of such extensions implemented in the current version of the model.

Our influence-spreading model has several main applications, including new definitions of out-centrality, in-centrality, betweenness centrality, and community influence measures. These measures are all based on different sums of the influence-spreading matrix elements, which allows for concurrent and consistent examination and analysis of network and community structures. The matrix elements of the influence-spreading matrix are probabilistic measures of influence spreading between all distinct nodes in the network. It is possible to measure the coherence of any subset of nodes in a network by adding up the influence-spreading matrix elements that connect nodes within that subset. When calculating the coherence of the entire network, all of the influence-matrix elements are added up. The influence-spreading matrix can be calculated based on different network models, making the method flexible for considering various network structures, interactions, and spreading processes.

The proposed model for measuring centrality in a network considers all possible paths between nodes, rather than just the shortest path as traditional methods do^{10,11}. This approach provides a more comprehensive understanding of the network structure. Additionally, the community influence method used for community detection identifies local maxima of a quality function, which is a novel idea not present in traditional community detection methods¹⁷. This approach enables the discovery of more detailed information about the network structure.

Here we propose an influence-spreading model that considers detailed-level node and link properties and all possible paths between nodes with path lengths less or equal to the maximum path length $L \leq L_{max}$. Our approach is a way of studying probabilistic spreading on complex networks. Next, we discuss several research questions related to the applications of the model and, more specifically, the use of the probability theory-based influence-spreading matrix. These include:

- How can we accurately define measures of generalised centrality and betweenness?
- How can we define a quality function to detect and analyse community structures?
- How can we identify communities, sub-communities, and their building blocks within the network?
- How can we incorporate various temporal distributions into the model? We extend our earlier method of using the Poisson distribution as a temporal distribution in the spreading process to include any other form of temporal distribution, including an empirical one. Different distributions can be included in the model by using probability distribution functions of time and path length. When time approaches infinity, the specific form of the temporal probability distribution does not affect the values of the influence-spreading matrix elements or derived results.

The main contributions of this study are:

- (1) We discuss the benefits of using shared methods and metrics in various applications for spreading models on complex networks, which is a new contribution to modelling principles. To achieve this, we introduce the influence-spreading matrix, a probability-based matrix that serves as a foundation for a range of derived quantities and applications. We can create this matrix using different simple or complex contagion models. For example, for static networks, we can use network connectivity models instead of spreading models. The influence-spreading matrix acts as a connector between network flow models and developing metrics or applications for various purposes.
- (2) To demonstrate the usefulness of our ideas, we summarise our influence-spreading model and its main applications. These applications are based on the influence-spreading matrix generated by the model.
- (3) We have defined novel measures based on the influence-spreading matrix. These measures include in-centrality, out-centrality, betweenness centrality, the cohesion of a set of nodes, and a quality function for

- community detection. Since these measures are calculated from the same matrix elements, they are consistent with each other. Our approach aims to avoid the mismatch discussed in the literature^{10,18,19} between standard centrality measures, network flow models, and community detection methods¹³.
- (4) We present a new approach to incorporate different forms of temporal distributions through a matrix as a function of path length and time. Technically, this is accomplished by using a file that allows the matrix to be replaced without altering other parts of the model or computer program. We have excluded some technical details and general topics that are beyond the scope of this study or have been presented in our earlier studies. These topics include simulation methods, demonstrations of different network sizes and structures, numerical benchmarks, and literature reviews of computational social networks.

Related work

It is challenging to prove theoretical results for the model outcome because it depends on the micro-structure of the network and the network flow process. This requires relatively strict constraints compared to the analysis of networks. Therefore, networks are usually classified according to standard quantities and investigated separately^{1,2}. Our study serves as a bridge between modelling detailed network structures and influence-spreading processes on social networks.

Complex contagion models have applications in many research areas of network science, such as influence spreading, opinion formation, and epidemic spreading. In the next subsection, we discuss related work of contagion models and interactions of higher order. Complex Contagion Models. One of the main applications is to define centrality and betweenness measures based on static or dynamic network models. In what follows we will summarise the standard centrality and betweenness measures mention some often-used community detection methods.

Contagion models and higher-order interactions

The formation of collective beliefs and the way individuals influence each other have been topics of interest in social sciences. Several models have been put forward attempting to explain the formation of opinions, group behaviour, and influence diffusion^{4,20–23}. Methods for modelling various dissemination processes on complex networks have been developed in network science^{24–26}. Mere pairwise interactions between individuals can sometimes fall short of characterising the propagation of social influence and shared dynamics, as underlined in²⁷. A more recent approach proposes a simplicial model to describe higher-order interactions²⁸.

Several mechanisms have been proposed to explain how influence spreads, including complex contagion^{29–31}. In⁹, complex contagion is described as a process in which an individual is exposed to multiple sources of influence, while simple contagion refers to the spreading of phenomena through a single exposure. Simple contagion is capable of accounting for certain epidemics and the dissemination of convincing gossip. In our approach, simple contagion is viewed as a fundamental propagation mechanism without breakthrough effects, whilst all other methods are considered complex contagions. Breakthrough influence happens when an already influenced individual gets re-influenced. Breakthrough effects are possible both in epidemic spreading and social contagion processes. This interpretation of contagion processes offers a more comprehensive understanding than what can be found in other references^{9,29,30}.

Compartmental models, which categorise a single population based on their health condition and age demographics, are typically employed to simulate the transmission of contagious diseases^{32,33}. These methodologies can likewise be utilised to influence spreading and social engagement on social platforms. For instance, the “Recovered” compartment can represent individuals who have let go of their convictions or viewpoints. In most cases, compartmental frameworks use discrete state variables, which are often found in simulation models. However, in analytical models, it's more common to use continuous state variables³⁴.

A literature overview³⁵ categorises research papers that examine non-drug-based approaches during the COVID-19 pandemic into four sectors: meta-population models, compartmental models, statistical models, and agent-based models. Meta-population models use a network of sub-populations that are linked via mobility. Statistical models evaluate the progress of the epidemic by deducing important parameters and actions from data. The spread patterns at a personal level are modelled by agent-based models³⁵.

When creating models for disease outbreaks and epidemics, Markov chains are often used⁵. The premise behind these models is that the propagation of diseases within networks can be portrayed as a random walk, where a walker transitions to a neighbouring node based on a probability linked to the current node. There are two types of random walk models: those with a single walker and those with multiple walkers. Models with multiple walkers are more accurate when describing opinion and epidemic spreading. However, closed formulas can only be obtained for specific symmetries or structures, and model variables must be calculated using agent-based methods and Monte Carlo simulations, as stated in the literature⁵.

In the realm of social science, agent-based models have been developed to simulate micro-level interactions. Recently, a probabilistic analysis of agent-based opinion formation models has been conducted to study the mechanisms behind opinion dynamics³⁶. The study mentions a long list of sociological and psychological concepts and references research in these areas such as attention, cognitive dissonance, polarity, leaders, stubbornness, tolerance, and emotions. The authors propose a technique based on probabilistic methodologies for the computational assessment of opinion formation models. This technique calculates the probability that the opinion distribution belongs to one of five qualitative categories³⁷. These categories are (1) Perfect consensus when the absolute majority chooses the same opinion; (2) Consensus, when most of the agents have a similar opinion; (3) Polarisation, when most of the agents are divided into two groups with distant opinions; (4) Clustering, when most of the agents can be grouped into two or more clusters; and (5) Disagreement of opinion, when there is no discernible tendency.

Mean-field reductions have been employed to investigate dynamics across a variety of fields, most notably in biological and neural networks³⁸. In research on the spread of rumours³⁹, a model was devised employing interacting Markov chains and mean-field equations to depict the dynamics in complex networks. In the mean-field theory, the average over an independent variable of the model is used to approximate the original model. The links between neighbouring nodes in a network facilitate pairwise influence in both Markovian and typical matrix algebraic network models. In these models, matrix algebra techniques like spectral clustering and eigenvector centrality have been utilised¹. Non-Markovian network models were created to incorporate higher-order interactions, where paths persist based on their origin⁴⁰.

In a study⁴¹, a general framework was proposed to link theory and empirical research in opinion formation models. The authors analysed their model using mean-field approximation and Monte Carlo simulations. The study used empirical data taken from an online social network. In⁴², mean-field approximation techniques are used to study how opinions spread and change in a society under the impact of influencer and media strategies. The study finds that influencers who take extreme positions tend to gain more followers. However, the paper also shows that optimal control strategies can be used by other influencers or media to mitigate such attempts and prevent further polarisation of opinions.

Political economists use game-theoretic approaches to develop equilibrium models of social interactions. This concept is studied in^{43,44} and covers a wide range of games with both local and global social interactions. Additionally,⁴⁵ examines the existence of pure strategies Nash equilibrium. The Nash equilibrium is a concept in game theory used to determine the best strategy for all players in non-cooperative games. The equilibrium occurs when each player has chosen a strategy, and no player has anything to gain by changing their strategy. On the other hand⁴⁶, introduces a game-theoretical generalisation of the landscape theory presented by Axelrod and Bennett⁴⁷. In a scenario where two groups are involved, each player rates their team based on the sum of their evaluations of group members.

Standard network centrality measures

In a network, nodes can be either central or peripheral in their position within the structure. The concept of centrality measures assumes the existence of a network process that flows through a network. Centrality measures can be seen as producing expected values for particular node results based on assumed models of the propagation process through a network¹⁹. This process identifies certain nodes as central to the network. However, it has been pointed out by various authors, including^{18,19}, that there is often a discrepancy between the actual network process and the assumed process model. Many of the standard centrality measures are not suitable for the types of flows that we are typically concerned with¹⁹.

Classical closeness centrality and betweenness centrality count only the shortest paths. Flow betweenness considers all the paths without self-intersection between nodes in the network⁴⁸. Eigenvector, and Katz centrality allow circular paths and multiple revisits on a node. There are two kinds of flows: conserving and non-conserving. In conserving flows, paths cannot be divided, while non-conserving flows allow propagation through multiple paths simultaneously^{11,19}.

Degree centrality is a straightforward measure of centrality. A node's degree is determined by the number of other nodes it is connected to. This is a local measure that does not consider the node's network position. Closeness centrality measures a node's centrality or influence on other nodes. On the other hand, betweenness centrality gauges the function of a node as an intermediary among other nodes in the network. Numerous versions of closeness and betweenness centrality measures exist, and the specific measure selected relies on the application. Other centrality metrics include those provided by mathematical models like eigenvector centrality. In the case of directed networks, definitions of out-centrality and in-centrality can also be applied. In its standard form, betweenness centrality is not defined for inward and outward directions¹.

Closeness centrality in its standard form is computed by determining the reciprocal sum of the shortest routes to other nodes within the network. The concept of distance is defined as the least amount of connections required to shift from one node to another, this is also equivalent to the length of the shortest path between the nodes. Betweenness centrality gets its definition from the frequency of a node appearing within the communication paths spanning between node pairs in the network. Nodes exhibiting high betweenness centrality values possess the ability to control or influence the information flow⁴⁹. Katz centrality enhances the definitions of degree and closeness centrality measures by taking not only the closest neighbours into account but also the total number of walks between a pair of actors^{1,11,50,51}.

It's worth noting that the standard centrality and betweenness measures are not always the best option for analysing network structures^{10,18}. This is because they are based on the shortest paths between a pair of nodes, which makes it challenging to interpret them in physical or probabilistic terms. Additionally, these measures differ from the ones proposed in this study in several ways. For instance, they do not take into account node or link weights, they are unable to model circular or self-avoiding interactions, or they do not enable the modelling of spreading processes on a network structure. Alternative centrality measures have been proposed in the literature but lack many of these properties. A recent study proposed a simulation algorithm called Path Information-based Centrality (NPIC)⁵². The algorithm considers various factors, such as node degree, shortest path distance, k-shell, adjustable parameters, and multiple neighbourhood levels. The authors compared their model's results with that of the well-known SIR (susceptible, infected, recovered) epidemic model³³. Furthermore, a centrality measure based on a generalisation of the SIR model has been proposed in⁵¹. Various issues and problems with centrality measures have been discussed in^{10,52-54}.

Community detection methods

A community is often described as a group of nodes in the network that are more closely connected than the nodes outside the group. Numerous methods and computer algorithms have been developed for modelling

community structure in complex networks. Graph partitioning involves dividing a network's nodes into non-overlapping groups of specified sizes while minimising the number of links between groups. A major challenge is that there is no widely accepted definition of what constitutes a community¹³, as different data and applications require different approaches. In many cases, the community concept is implicit in the method or algorithm used. As new algorithms are developed, the number of definitions of a community will continue to expand. Community detection methods typically involve optimising a quality function, with modularity being a common choice. The mathematical definition of modularity can be found, for example, in^{1,2}.

Infomap and the Louvain algorithm are two fast methods for identifying communities in very large networks. They have become popular due to their effectiveness in optimising quality functions. The Louvain algorithm uses modularity to optimise quality, while Infomap uses an entropy-based measure. The Louvain algorithm optimises modularity through local changes and builds larger communities by aggregating modules. Infomap condenses the data relating to a random traveller exploring the diagram¹.

Recently, community-aware methods have been proposed as an alternative to classical centrality measures^{55–57}. These methods use a network's community structure to identify influential nodes in complex networks. One approach⁵⁷ involves determining a structural central node in each community and partitioning other nodes into different communities based on a multi-source diffusion and majority voting process. Another method⁵⁵ ranks nodes in a sequential order linked to the community size, selecting nodes across all regions of the network. The authors of these studies highlight the importance of incorporating community structure information in defining centrality measures to better rank influential nodes. The relationship between classical and community-aware centrality measures has been studied in⁵⁶, and the results show that they generally exhibit low to medium correlation values. The authors state that integrating knowledge about the network community structure brings a new perspective of node influence, paving the way for the development of effective community-aware centrality measures.

Clustering of graphs based on stochastic models, or stochastic block models, are commonly used for detecting community structure in networks and generating synthetic benchmark networks⁵⁸. Some community detection methods can also identify hierarchical and overlapping sub-communities in complex networks. A recent study¹⁴ has presented an information-theoretic method for discovering building blocks of network nodes that are typically found in the same community.

Typically used for detecting community structures in networks and creating synthetic benchmark networks, graph clustering is often based on stochastic or stochastic block models⁵⁸. Some community detection methods can also identify hierarchical and overlapping sub-communities within complex networks. A recent research¹⁴ has illustrated an information-theoretic method for identifying building blocks of nodes that are often found in the same community.

Model

In this section, we demonstrate the modelling of the spreading process by combining different paths of the network structure from a source node to a target node. Unique in the model is that it describes node and link level spreading in the network structure, and in addition, it considers temporal dependency on individual paths. In the next subsection, we explain the mathematical basis and main ideas of the model. Detailed algorithms have been proposed in our earlier studies⁶. Unique in our model is the method used to process various spreading paths to achieve their combined effects. This process is outlined in Algorithm 1⁶. To efficiently calculate effects in larger networks, we have developed Algorithm 2 detailed in⁶.

The fundamental concepts in modelling influence spreading on social networks include network structure, link and node weights, paths, and spreading rules. Network structure is defined by directed links between adjacent nodes, and each link is given a numerical value of spreading probability. Paths are lists of nodes in the network structure^{1,2}.

We propose a model that calculates time-dependent spreading probabilities between individual nodes at the microscopic level of the network structure. The model takes into account directed connections with individual weights for nodes and links. When influence starts at a source node and spreads through a path, its probability is determined by two main factors: the first factor describes the temporal development of the spreading, while the second factor describes the activity of links and nodes along the path.

As time passes, the first term that describes temporal development approaches one. This means that the probability of influence spreading through a path is the product of all node weights and link weights along that path. For a finite time, the temporal development is described by the complementary cumulative distribution function which is called a survival function. In this case, the survival function gives the probability that the influence has spread at least the amount of the path length or the number of links on the path.

The temporal development on a path is expressed as a survival function $\mathcal{P}_L(T)$ of a probability distribution function $\mathcal{D}_L(T)$ as

$$\mathcal{P}_L(T) = 1 - \mathcal{D}_L(T). \quad (1)$$

For example, if the spreading process obeys the Poisson distribution, we have

$$\mathcal{D}_L(T) = \left(\sum_{z=0}^{L-1} \frac{(\lambda T)^z}{z!} \right) e^{-zT}, \quad (2)$$

where λ is the intensity parameter of the Poisson distribution, T is time and L is the number of links on the path. One can use different distributions by inputting the corresponding probability distribution function $\mathcal{P}_L(T)$ into Eq. 1.

The second factor consisting of link and node weights is expressed as

$$P(\mathcal{L}) = \prod_{(u,v) \in E_{\mathcal{L}}} w_{(u,v)} \cdot \prod_{v \in \mathcal{L}} w_v \tag{3}$$

where $w_{(u,v)}$ and w_v denote the spreading probabilities through the link from node u to node v and through node v , respectively. The set of links on the path \mathcal{L} is denoted $E_{\mathcal{L}}$.

Because we assume that the temporal development is independent of the link and node weights, the probability of influence spreading via path \mathcal{L} at time T is expressed as the product of the survival function and the link and node weight terms as

$$p(\mathcal{L}, T) = \mathcal{P}_L(T)P(\mathcal{L}). \tag{4}$$

In our primary model, we assume that the spreading probabilities through a node are independent of the node states. We have previously presented a simulation model for breakthrough effects¹⁵, which considers node states. Later, in our demonstrations, we assume that the spreading process starts with probability one from the source node. Considering the starting probabilities can be included in the calculations simply by using the relevant node weights for each starting node of the paths. The third assumption is that the network structure is static in the sense that nodes or links are not removed or added during the spreading process. Considering these kinds of events in the model is a subject of future work because they are not included in our current algorithms.

In the model, influence spreads through multiple paths in the network structure at the same time. The method of spread is non-conserved, making it suitable for describing how opinions and epidemics spread. It is also assumed that the probabilities of spreading through alternative paths are independent.

Next, we explain how the algorithm works (see [Supplementary information](#) online for more details). We denote the maximum path length for computing the probabilities as L_{max} . Model parameter L_{max} is needed to limit the use of computing time. The parameter can also be used to study spreading probabilities on different ranges in the network structure. The method is based on computing the influence probabilities starting from the maximum path length and proceeding in the reverse order of path lengths until all possible paths are considered. This approach has the advantage that lists of different paths need not be saved in memory.

We compute the spreading probabilities from source nodes inside the range of L_{max} to the target node e . The calculation is performed for a fixed value of time T . We are often interested in the limiting case where $T \rightarrow \infty$ when node states reach stable values. We do not indicate time T in the following formulas to simplify notations. Also, we do not indicate the Poisson parameter λ because the formulas are expressed in a general form for any temporal distribution.

The idea of the algorithm is illustrated with an example in Fig. 1. For path length L We define a new variable $p(u, v, L)$ by using the known temporal factor $\mathcal{P}_L(T)$ and the known node and link weights of the current iteration step. For example, for $L = L_{max} - 1$ from a neighbouring node n to target node e , $p(n, e, L_{max} - 1) = w_e w_{n,e} \mathcal{P}(L_{max})$ where w_e is the node weight and $w_{n,e}$ is the link weight. Next, we iterate path length L from $L_{max} - 1$ to 0.

If there is a link from node m to node n , we have $p(m, n, L_{max} - 2) = w_n w_{m,n} w_e w_{n,e} \mathcal{P}(L_{max})$. If a link exists also from node m to node e , we have $p(m, e, L_{max} - 1) = w_e w_{m,e} \mathcal{P}(L_{max} - 1)$. Because these two contributions have a common prefix of length $L_{max} - 2$, they are not independent. The rules of probability theory are applicable, and we use the non-mutually exclusive rule to combine the effects of the two paths. The combined effect is

$$p(m, e, L_{max} - 2) = p(m, n, L_{max} - 2) + p(m, e, L_{max} - 1) - \frac{p(m, n, L_{max} - 2)p(m, e, L_{max} - 1)}{\mathcal{P}(L_{max} - 2)} \tag{5}$$

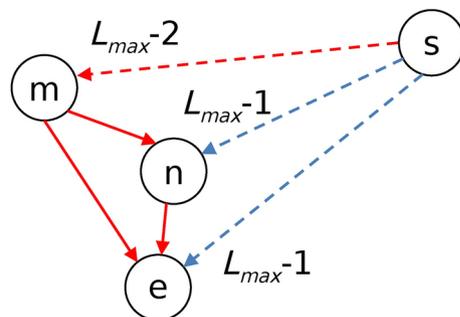


Fig. 1. Connections in the example to illustrate the combination of paths in the influence-spreading model⁶.

If there are more paths to be combined for the same value of L , they can be calculated similarly. These paths can be processed in any order as long as the paths are combined in a descending order based on the lengths of the longest common prefixes of the paths. After repeating the procedure from $L = L_{max}$ to $L = 0$ we get $C(s, e) = p(s, e, 0)$, where $C(s, e)$ is the probability of influence spreading from source node s to target node e . We name matrix $C(s, e); s = 1, \dots, N; e = 1, \dots, N$ influence-spreading matrix.

Paths are processed iteratively by repeatedly applying the method of combining paths to calculate more and more paths. The reverse calculation makes sure that all path pairs are merged at the last possible node, as dictated by the model. The outcome is the cumulative probability of influence disseminating from one node to another via all viable pathways with $L \leq L_{max}$. No limitations are imposed for breakthrough spreading through nodes, which means that all loops and recurrent transitions are allowed during the spreading process.

In our model, the probability of reaching the target node applies to the initial impact of alternative paths of influence from the source node. We do not factor in the secondary effects of the target node on itself, even though we consider the circular effects of all other nodes on the path. In this study, we do not make a clear difference between opinions, beliefs, behaviours or attitudes. However, model parameter values and interpretations of the model results may differ in these cases⁴.

Applications of the model

In this section, we present some examples of applying the influence-spreading matrix approach in the analysis of the complex network and its elements. In the next subsection, we discuss the centrality and betweenness measures that are basic concepts in the network analyses. In our demonstrations, we use the network structure of Fig. 2a. The building blocks of the same network discovered by an information theoretical community detection method can be found in¹⁴. These building blocks can be compared with the corresponding results of our model. Nodes in the network represent characters, and edges connect any pair of characters that appear in the same chapter of Victor Hugo's novel *Les Misérables*. In the subsequent sections, we discuss different spreading processes, network resilience, temporal spreading distributions, and effects depending on the path length.

Centrality and betweenness measures

We will now discuss three metrics used to determine the significance of a node in a network: out-centrality, in-centrality, and betweenness centrality. Let V denote the subset of nodes in the network G . The influence-spreading matrix $C(s, e)$ is defined in the Model Section. The out-centrality of node s is calculated as

$$C^{(out)}(s) = \sum_{\substack{e \in V \\ s \neq e}} C(s, e) \quad (6)$$

and the in-centrality of node e as

$$C^{(in)}(e) = \sum_{\substack{s \in V \\ e \neq s}} C(s, e). \quad (7)$$

These metrics have a straightforward interpretation as expected values. The out-centrality is the expected number of nodes it can influence, while the in-centrality of a node is the expected number of nodes that can influence it. For instance, the out-centrality can be used to anticipate the number of people who might be influenced by

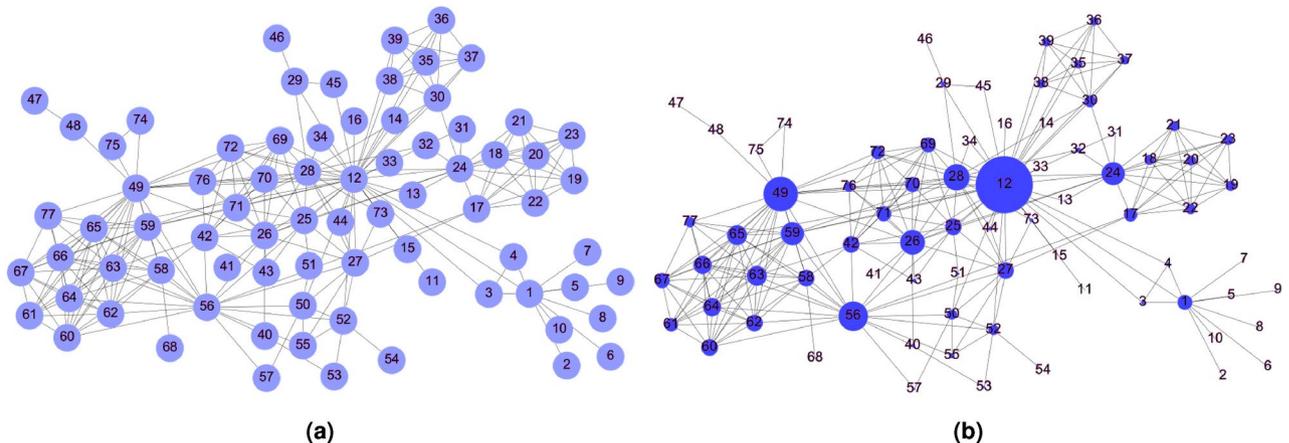


Fig. 2. The network structure of the *Les Misérables* network used in the demonstrations of this study. **(a)** The network structure without scaling and **(b)** the network structure with node sizes weighted by their node degrees.

a particular node, while the in-centrality can be used to assess an individual's vulnerability to misinformation. Figures 3 and 4 show out-centrality and in-centrality values of the nodes in the Les Misérables network of Fig. 2 with link weights $w = 0.1$ and $w = 0.5$, respectively. These values represent low and moderate link strengths and provide sufficient examples to illustrate the model's characteristics. High link weight values close to $w = 1$ provide more stable results due to the saturation effects, and they are rare in social network applications. To highlight the differences more clearly between the out-centrality and in-centrality results with link weights $w = 0.1$ and $w = 0.5$ we have used the Gephi network analysis and manipulation tool⁵⁹ in Figs. 3b,c, 4b,c.

For low link weights, the spreading probabilities are lower than for higher link weights, as we can see in Figs. 3 and 4. More interesting is that central nodes are more emphasised in the out-centrality than in the in-centrality results. These effects are even more striking for low link weights in Fig. 3. The explanation for both observations is that the neighbouring nodes are more dominant for the spreading process to get a good start in these situations. From Fig. 3, we can identify the most central nodes 12, 49, 56 and 59. It is also possible that the rankings of nodes change with different link weights. For example, node 1 has a higher ranking for higher link weights $w = 0.5$ than for lower link weights $w = 0.1$. Moreover, when the link weight is 0.1, the relative in-centrality value of node 1 is higher than the out-centrality value. However, when the link weight is 0.5, the opposite is true.

The out-centrality and in-centrality of a node in Eqs. 6 and 7 represent average values of influence-spreading matrix element values in the network structure. These centrality measures can also be expressed in percentages. It is also possible to calculate detail-level quantities for a node or a set of nodes. Figure 5 provides an example for node $n = 12$. Node 12 has a degree of 36 and is the most central node in the network when measured by degree. Other central nodes, in the descending order of degree (degrees are indicated in the parentheses), are nodes $n = 49$ (22), 56 (19), 28 (17), 26 (16), 24 (15), and 59 (15). To visualise the network structure based on node degrees, Fig. 2b displays the same network as Fig. 2a, but with node sizes weighted by their degrees.

By comparing Fig. 5b,c with Fig. 3b,c, several observations can be made. The numerical values of out-centrality and in-centrality in Fig. 5a, expressed as the number of influenced or influencing nodes, are lower

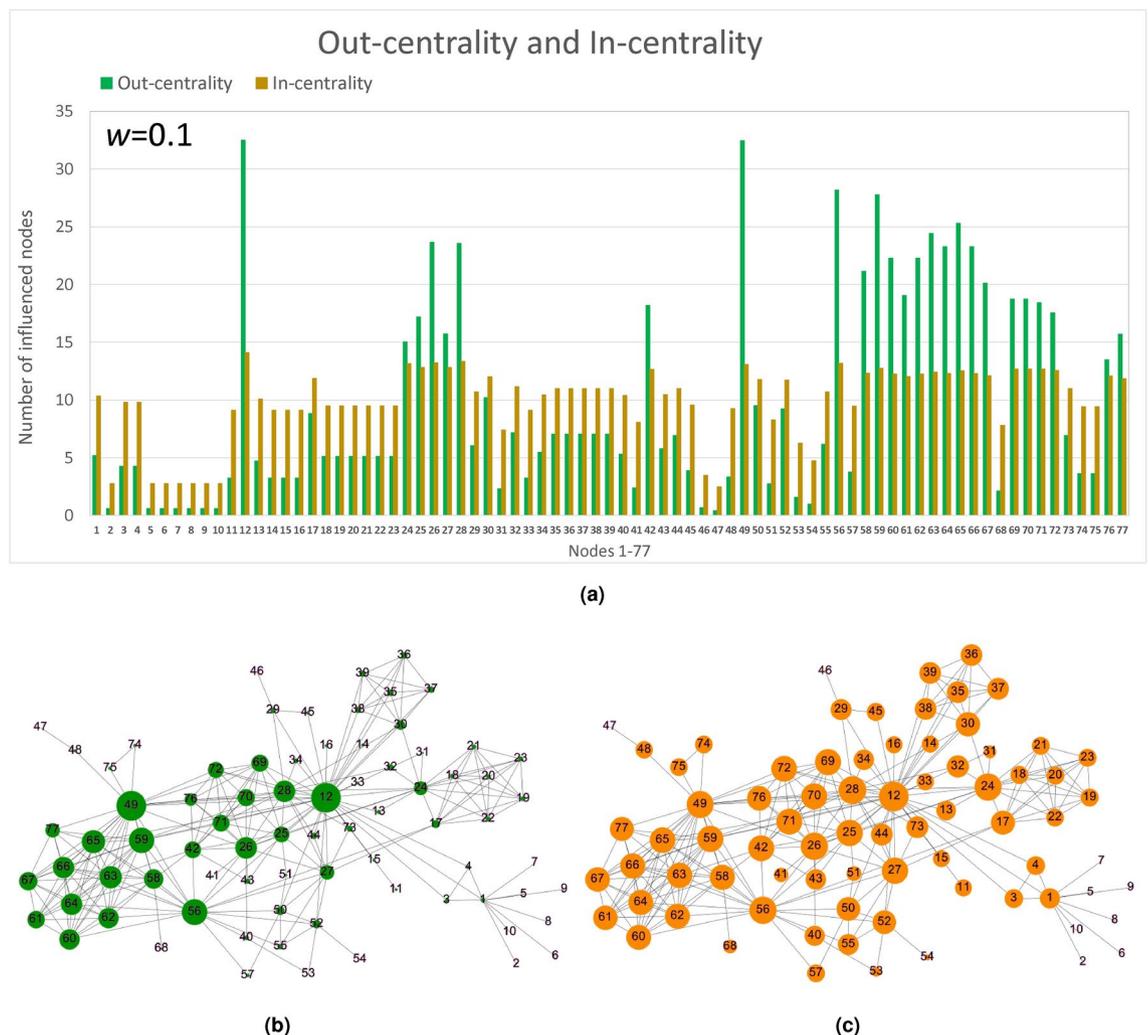


Fig. 3. Out- and in-centrality for low link weights. (b) and (c) Illustrate the out- and in-centrality values of (a) where node sizes are weighted by their corresponding centrality values.

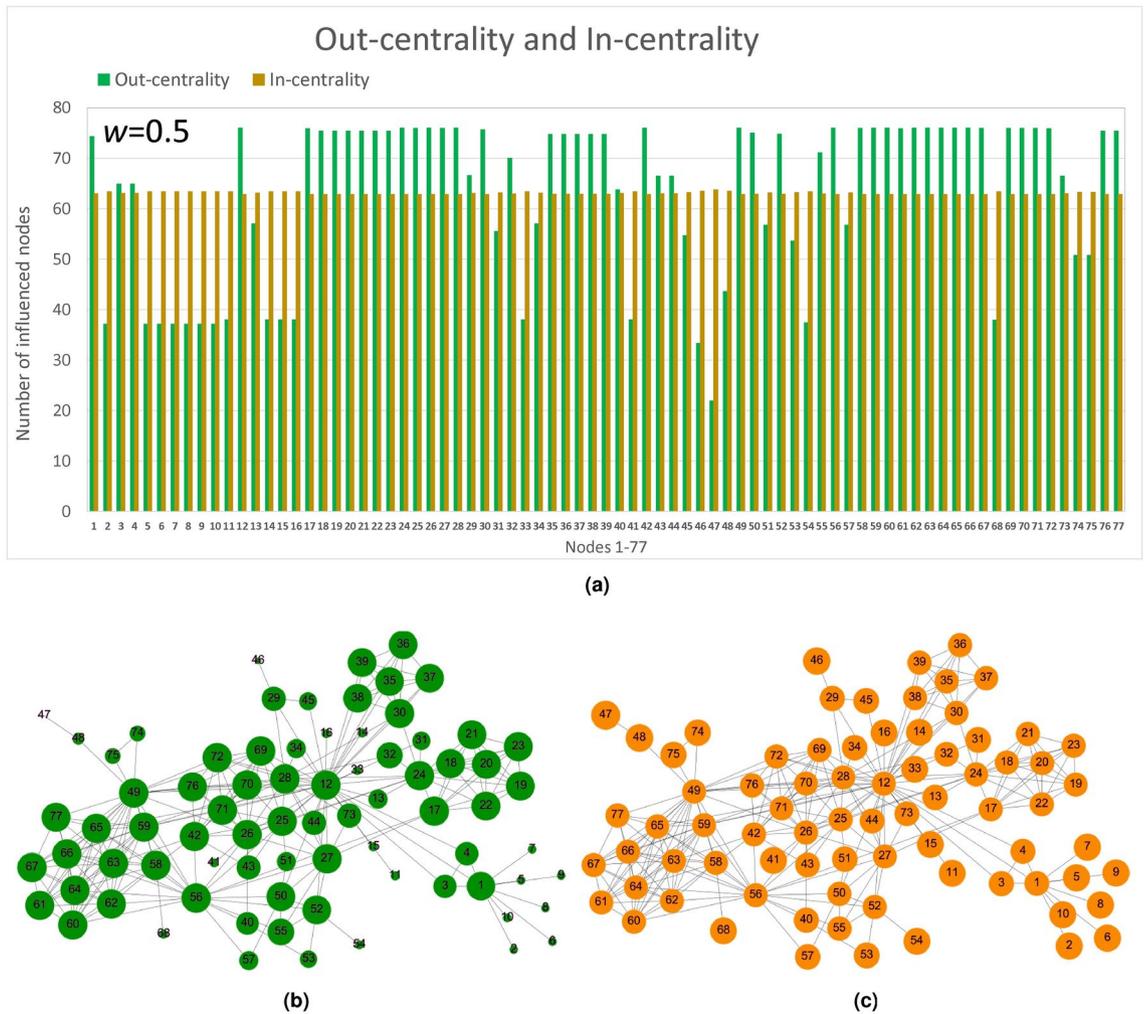


Fig. 4. Out- and in-centrality for high link weights. (b) and (c) Illustrate the out- and in-centrality values of a where node sizes are weighted by their corresponding centrality values.

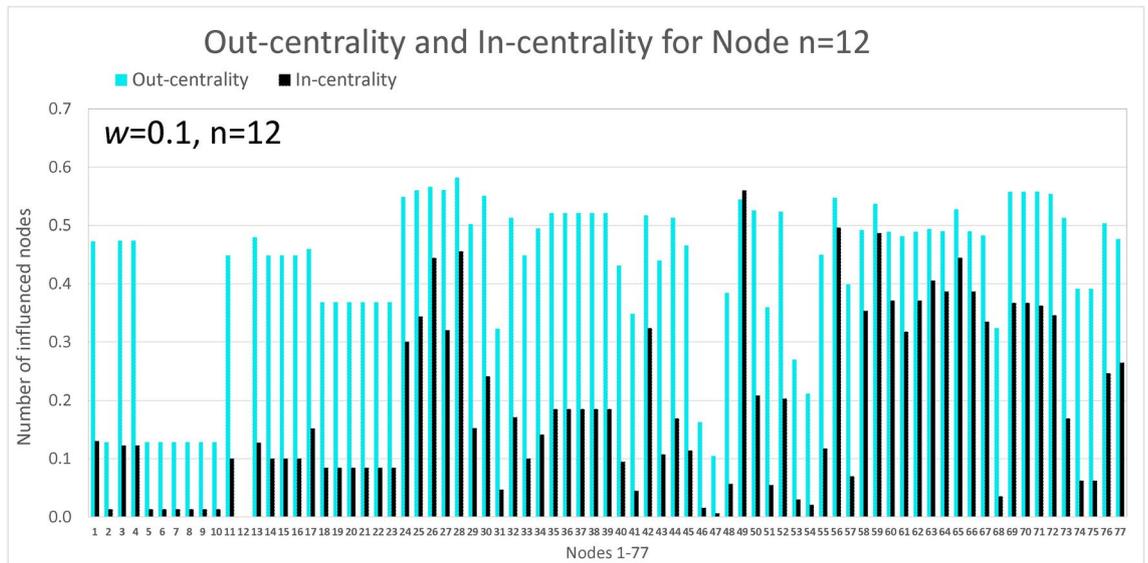
compared to Fig. 3a. This is because Fig. 5a shows the breakdown of out-centrality and in-centrality values of node $n = 12$ for each of the other 76 nodes in the network. The sums of node-specific out-centrality and in-centrality values of Fig. 5a are $C^{(out)}(s = 12) = 32.5$ and $C^{(in)}(e = 12) = 14.1$, respectively, as calculated from Eqs. 6 and 7. These values represent the out-centrality and in-centrality of node $n = 12$ in Fig. 3a.

In the influence-spreading model and other flow-based network models⁵³, a node’s influence is not limited to its direct connections but also affects the influence between distant nodes in the network as it transmits influence. This amplifies the effects of central nodes, especially if their neighbours in the network also have many connections. Out-centrality and in-centrality are two measures that determine the centrality of nodes in a network based on their ability to influence and be influenced by other nodes. These measures, as proposed in Eqs. 6 and 7, differ from the simplest centrality measure of degree, which only considers the local structure of the network. In-degree and out-degree can also be defined for incoming and outgoing connections, respectively. However, in networks with bidirectional connections, both measures have the same values. It is important to note that in our influence-spreading model, out-centrality and in-centrality values are different also in the bidirectional case, except in some symmetric network structures.

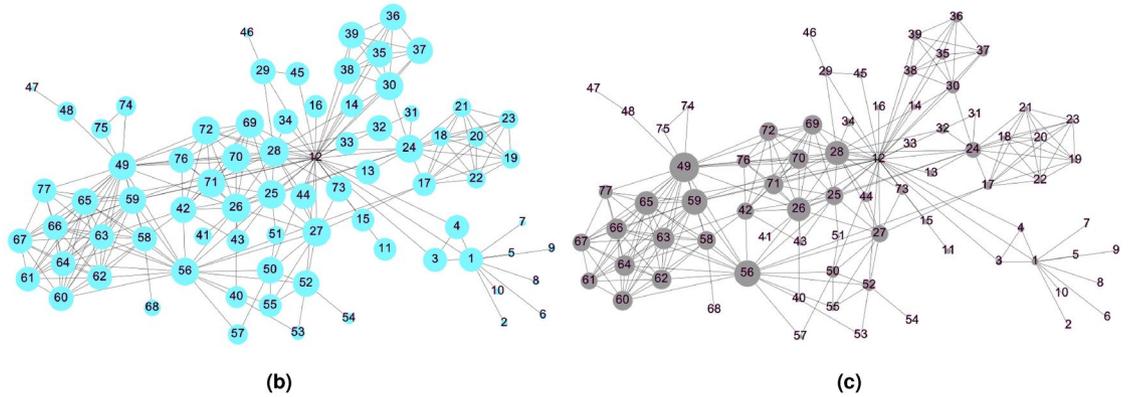
Figure 5b shows that node 12 has a similar influence on most other nodes. However, in Fig. 5c, the influence of other nodes on node 12 looks like the average out-centrality results of the entire network in Fig. 3b. Node 12 plays a dominant role in the network, as evidenced by the out-centrality values of the entire network being alike to the in-centrality values of node 12. Node 12 serves as an example of how node-specific results can differ significantly from the average results of the network.

We define betweenness centrality b_s for node s as the relative difference in the cohesion of the network when node s is removed from the network. The cohesion \mathcal{C} of network V is defined as

$$\mathcal{C} = \sum_{s, e \in V} C(s, e) \tag{8}$$



(a)



(b)

(c)

Fig. 5. Out- and in-centrality for one node ($n = 12$). (b) and (c) Illustrate the out- and in-centrality values of (a) where node sizes are weighted by their corresponding centrality values.

and the cohesion \mathcal{C}_s of network V with node s removed as

$$\mathcal{C}_s = \sum_{\substack{s, e \in V - \{s\} \\ s \neq e}} C(s, e). \tag{9}$$

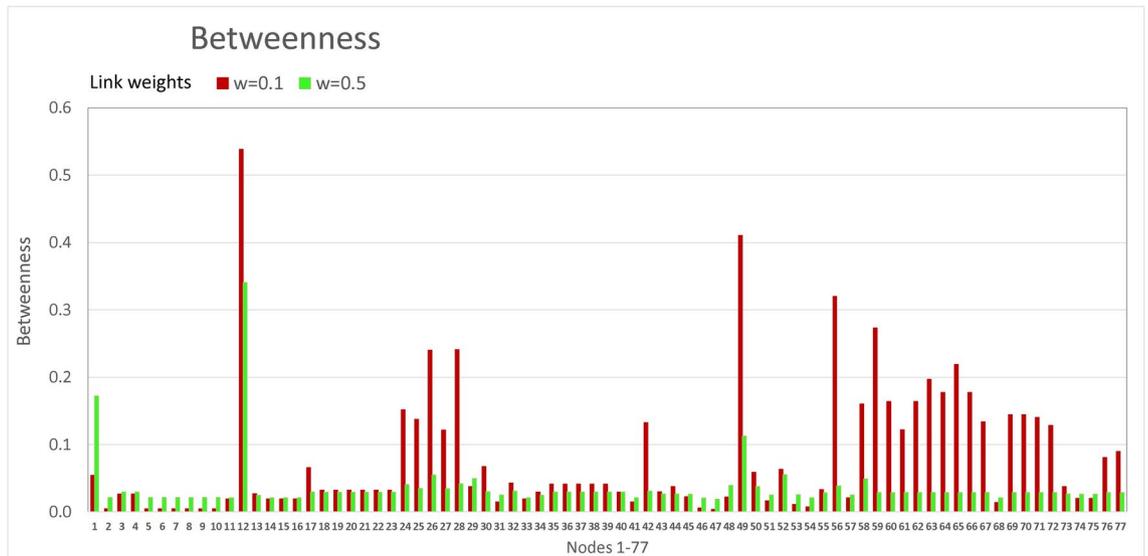
The betweenness centrality of node 12 is defined as

$$b_s = \frac{\mathcal{C} - \mathcal{C}_s}{\mathcal{C}}. \tag{10}$$

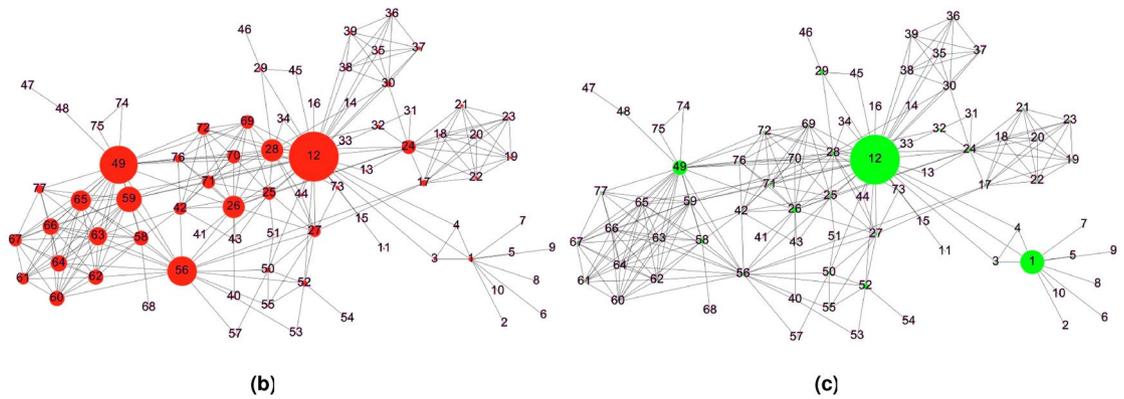
The betweenness centrality value of multiple nodes can be calculated similarly to that of a single node. This metric can help identify crucial nodes that facilitate the flow of information or influence between parts of a network structure.

Figure 6a shows the betweenness centrality values of the nodes in the Les Misérables network of Fig. 2a with link weights $w = 0.1$ and $w = 0.5$. To highlight the differences more clearly between the betweenness results of link weights $w = 0.1$ and $w = 0.5$ we have used the Gephi network analysis tool in Fig. 6b,c. For low link weights nodes 12, 49, 56, and 59 have high betweenness values when compared with other nodes. Node 1 has a low betweenness value for low link values. For higher link weights the betweenness values of nodes 1 and 12 are relatively high.

The results of betweenness centrality share some similarities with the out-centrality and in-centrality results in Figs. 3 and 4. In our current model, only one betweenness measure is defined for each node, which is a stable measure as a function of the link weight. However, the relative values and node rankings can change with different link weights. For instance, node 1 has a higher betweenness for the link weight of $w = 0.5$ than for the



(a)



(b)

(c)

Fig. 6. Betweenness for low and high link weights. (b) and (c) Illustrate the betweenness values of (a) where node sizes are weighted by their corresponding betweenness values.

weight of $w = 0.1$, which is contrary to most other nodes in the network. The reason for this is that node 1 is in a gate-keeping role in the network and it can disconnect a great number of nodes (7 nodes) in the network. As a comparison node 49 disconnects only four nodes, and node 1 itself has much better connectivity via its neighbouring nodes to the network. The betweenness value of node 49 is relatively lower for high link weights than for low link weights.

Community detection

In this study, the method for detecting communities is based on finding local maximum values of the community influence measure in Eq. 11 computed from the influence-spreading matrix elements $C(s, e)$; $s, e = 1, \dots, N^8$.

$$P = \sum_{s,e \in V} C(s, e)(T) + \sum_{s,e \in (G-V)} C(s, e)(T) \tag{11}$$

The first sum involves pairs of nodes, represented by s and e , that belong to a subset V of a network G , while the second sum involves pairs of nodes in the remaining network partition (denoted by $G - V$). Equation 11 includes rows and columns corresponding to node pairs within and outside the community. This method allows for a detailed network system model by including all possible paths between two nodes.

In Eq. 11, matrix $C(s, e)$ represents the influence-spreading matrix, which captures the influence of nodes on each other. For static networks, a different network model, such as a connectivity model⁸, can be employed to produce a time-independent connectivity matrix to replace the influence-spreading matrix.

Equation 11 is designed to analyse processes at different time scales during the spreading process. It can also be used to examine static network structures with a time of spreading approaching infinity $T \rightarrow \infty$. The model can utilise individual weights for each node and link in the network, allowing for a study of the network with different interactions between nodes, resulting in more precise results when such information is available.

Our optimisation strategy assumes that the network can be divided into two communities. Despite this, the model may offer various solutions for the local maximum of Eq. 11. One general description of a community is a network component whose nodes are more tightly connected than nodes outside of it^{1,2}. We propose that the formation of a community has a specific reason to differentiate it from the other part of the network, whose tightly connected nodes are also regarded as a community. The connections between the two divisions must be relatively scarcer or weaker than the connections within the two divisions. This kind of scenario is a requirement for local maxima of the quality function to emerge.

Communities that score highly based on quality function P in Eq. 11 are viewed as potential nominees for the division of the initial community in practical scenarios. However, it is worth noting that this might not be the most likely solution for the formation process of the community. A statistical indicator to portray the likelihood of the community's formation is defined in⁸.

The quality function for community detection in Eq. 11 and the out- and in-centrality measures in Eqs. 6 and 7 are based on the same influence matrix $C(s, e); s, e = 1, \dots, N$. Therefore, these measures are inherently consistent with each other. As a result, it may be unnecessary to define specific community-aware centrality measures, as proposed in^{56,57}. Instead, we recommend considering our proposed out-centrality, in-centrality, and betweenness measures.

In our methodology, the link weights in the influence-spreading model directly affect the number of detected sub-communities in the community detection model. The higher the link weights, the fewer sub-communities are emerging. Link weights represent the probability of influence between two adjacent nodes in the network, and the quality function of community detection is also interpreted as a probability. We can also make estimates about reasonable ranges of the parameter values. Based on our community detection method, we can conclude that sub-communities emerge when link weights are relatively low in real-life situations.^{7,8} On the other hand, higher link values exist in situations where the cohesion of the social network is high, and no sub-communities emerge. This conclusion is based on the detailed-level modelling of interactions in social networks and the assumption that the model adequately describes real life. In addition, quantitative measures allow intuitive evaluations and judgement of the numerical values of their magnitudes.

The investigation of the Les Misérables network in¹⁴ uses general information theory instead of a specific community detection method proposed in the literature. The results of our model in Fig. 7 and the information-theoretic findings¹⁴ agree with each other. The normalised mutual information measure is commonly used as a reference when new community detection methods are introduced and because it can be regarded as a measure that has theoretical grounding although it has been found to be biased⁶⁰.

As previously mentioned, there is no clear definition of what constitutes a good community. Numerous community detection methods have been proposed in the literature, many of which rely solely on local information of the network structure, or a combination of metrics^{13,61}. However, methods based on modularity optimisation¹², such as the Louvain algorithm, tend to prioritise larger networks and fail to identify smaller sub-communities^{61,62}. Our method is based on identifying local maxima of the quality function in Eq. 11, which allows us to detect smaller sub-communities and building blocks of communities⁸. While our method provides an alternative approach to addressing the consistency problem and resolution bias, it may not completely resolve the challenge of determining what qualifies as a community in all contexts.

Spreading processes

In the literature, spreading processes have been categorised into two classes: simple contagion and complex contagion^{9,31,63}. There are no universally agreed mathematical definitions of the concepts of simple and complex contagion mechanisms. In this study, we consider simple contagion using the network flow model based on self-avoiding paths, and complex contagion in all other cases. Simple contagion models are useful for modelling

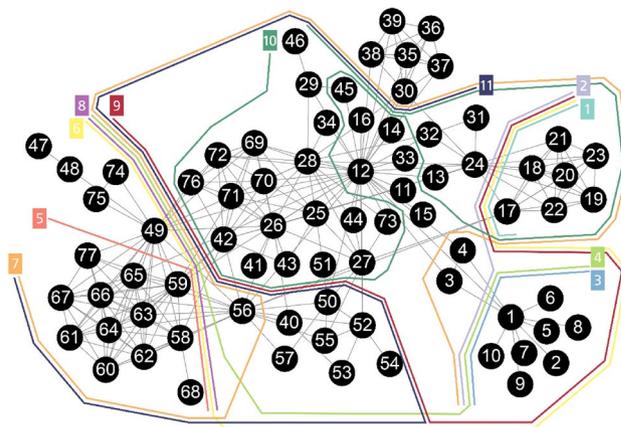


Fig. 7. Overlapping and hierarchical communities. Community structure detected by our method with Eq. 11 from the Les Misérables network of Fig. 2a. The most important divisions are indicated by the eleven lines in the figure^{8,17}.

information transfer and epidemic spreading with immunisation while complex contagion models are useful for modelling rumour and behaviour spreading.

Social interactions consist of casual exchanges and formal knowledge sharing, and the spread of diseases typically involves various transmission mechanisms and some immunity acquisition post-infection recovery¹⁵. One approach to model these network processes is through breakthrough mechanisms¹⁵.

Influence spreading and opinion formation

We have proposed a model that allows all individuals in a social network to interact, including circular influence spreading through loops and recurrent interactions between neighbouring nodes in the network structure. In our basic model, no additional or restrictive rules are used for the spreading process. Technically, we can say that the model allows full breakthrough effects through nodes of the network irrespective of the node states.

When compared to simple contagion models, circular and recurrent effects can alter the importance of nodes. When link values are low, only the links between the source node and its neighbouring nodes are important. Hence, when link weights are small, the simple contagion model with self-avoiding paths and the complex contagion model can yield similar results. In the most simple contagion models, node rankings remain unchanged as link weights increase. However, this is not generally the case with our complex contagion model^{7,8,15}.

Epidemic spreading

We have used our model to investigate how epidemics spread within organised structures. The COVID-19 virus is primarily transmitted through respiratory droplets and aerosols from infected individuals, but it can also be spread through physical contact. Since the virus spreads through social interactions and gatherings, people with strong social ties tend to meet more frequently indoors, increasing the likelihood of transmission. Our model uses weights to represent the probability of infection transmission from one person to another during a disease's contagious period. It also takes into account the frequent gatherings of close friends, which can lead to repeated and circular influence in the social network^{15,16}. Additionally, the model considers the potential for breakthrough infections originating from people who have previously recovered from the infection and may still be capable of spreading the virus¹⁵.

We have determined the impact of reducing different types of contacts within an organisation, as well as compared various preventive measures and their combinations to the scenario where no action is taken. This information is useful in identifying the most effective preventive measures to contain the spread of an epidemic¹⁶.

Measures such as out-centrality, in-centrality, and betweenness have been employed to comprehend different facets of epidemic spreading. Community immunity, also referred to as herd immunity, is established when a significant proportion of the community is immunised to a point where the basic reproduction number is one or less ($r_0 \leq 1$). In our model, the relationship between link weights and the basic reproduction number has been presented in¹⁶.

The experimental design for studying epidemic outbreaks has some limitations. The model only considers two states for individual nodes in the network: susceptible and infected. While this is useful for comparing different preventive measures, it does not fully capture the entire process of epidemic spreading, which involves exposed and recovered node states. Individual immunity variations can be computed through iterative calculations, but a more comprehensive solution would require further model development. To realistically represent varying individual immunity during the incubation and infection periods, the model should incorporate multi-variable modelling of time and state dependencies at the node level. The current model only considers temporal dependence as a function of the path length.

Information spreading

Our definition of the simple contagion model includes self-avoiding paths and disallows recurrent and circular effects. This model, which prohibits breakthrough events, can be applied to simulate information spreading within a network structure. The reason for this is that no repeated transfer is needed and events end at a node where circular or repeated events would start.

In our study⁶⁴, two models for information spreading and communication network connectivity were compared in terms of calculation methods and results. Despite being designed for different applications, the two models produced mathematically equivalent results when the maximum spreading path length, L_{max} , was sufficiently high.

Network resilience

The classical model for network connectivity aims to explain the dependability of connectivity in service networks. According to the model, information spreads through a network when one individual comes into contact with another, with the spread continuing through subsequent contacts. In information networks, connectivity refers to information spreading through network nodes and connections, with connections existing wherever information can be transferred.

Our case study⁶⁵ examined a multi-layer communication network system consisting of a backbone network, two access networks, and a satellite system. The model treated the system as a single network, enabling the computation of resilience between the elements or network structures within it.

This model has practical applications such as analysing the resilience of individual networks, examining links across networks, and comparing network topological structures. Furthermore, it can be utilised in studying the resilience of any element, connection, node, or group of elements within the network. This information can assist in designing and constructing more resilient and more efficient system arrangements with better quality of services.

Temporal distributions

In Fig. 3a we have presented the results for out-centrality and in-centrality measures with link weights $w = 0.1$ when time approaches infinity $T \rightarrow \infty$. Figure 4a shows the corresponding results with link weights $w = 0.5$. In this section, we break down these results into different contributions for finite time values. Our model allows for the analysis of average out-centrality and in-centrality values, as well as detail-level spreading probabilities between individual nodes in a network structure.

Different models, such as Poisson distribution or email passing distributions, can be used to study the effects of different temporal distributions. Empirical temporal distributions can also be inputted into the model. Different distributions can be included in the model simply by implementing a file where the distribution is given as a function of time and path length. An example of the file format is provided in Supplementary Table S1 online. No changes need to be made to other parts of the model. When time approaches infinity $T \rightarrow \infty$, the specific form of the probability distribution does not affect the values of the influence-spreading matrix or centrality measures.

In Figs. 8 and 9, we illustrate the temporal dependence of the out-centrality and in-centrality values of the spreading process in the example network of Fig. 2a with the link weight values of $w = 0.1$. Corresponding results for the link weight values of $w = 0.5$ are shown in Figs. 10 and 11. Poisson distribution with the intensity parameter value $\lambda = 1$ in Eq. 2 is used for describing time dependence along a path in the network structure. Spreading as a function of time is depicted with stacked histograms where time values are displayed with the colours indicated by legends in the figures. Figures 8, 9, 10 and 11 can be compared with the corresponding Figs. 3 and 4 where the out-centrality values (in-centrality) of final node states for high values of time $T \rightarrow \infty$ are presented.

In Fig. 10, it can be observed that the spread with high link weights ($w = 0.5$) occurs faster than with low link weights ($w = 0.1$) in Fig. 8. The more isolated nodes, such as nodes 2, 5 – 10, are influenced later (between times $T = 3$ and $T = 4$, as indicated by the blue shades in the figure) compared to well-connected nodes in the network structure. This sub-community, which includes node 1 and ranks third, merges with nodes 3 and 4 to form the fourth-ranked community in Fig. 7. These sub-communities are an example of overlapping communities. They can be compared with the first-ranked community of nodes 17 – 21 in Fig. 2a, which is better connected to other parts of the network through nodes 17 and 24.

These examples illustrate how influence spreads at different times of the spreading process, from individual nodes to sub-communities. The results are based on aggregated values from the directed spreading probabilities $C(s, e)$, $s, e = 1, \dots, N$ between the N individual nodes in the network. Observing Figs. 8 and 9, we can discern significant differences when a node is exerting influence on other nodes in the network and when other nodes are influencing a node in the network. This model feature can be used to analyse the temporal development of opinion spreading and community formation in social networks.

Figures 10 and 11 demonstrate the effects of higher link weights $w = 0.5$. By comparing these results with Figs. 8 and 9, we can observe how the out-centrality and in-centrality measures evolve differently over time. Such analyses reveal temporal fine structure in the long-term results calculated for high values of time T in Fig. 4. Figures 8 and 10 show that the out-centrality values for low time values with high link weights have similarities with the out-centrality values for high time values with low link weights. Figures 9 and 11 show the

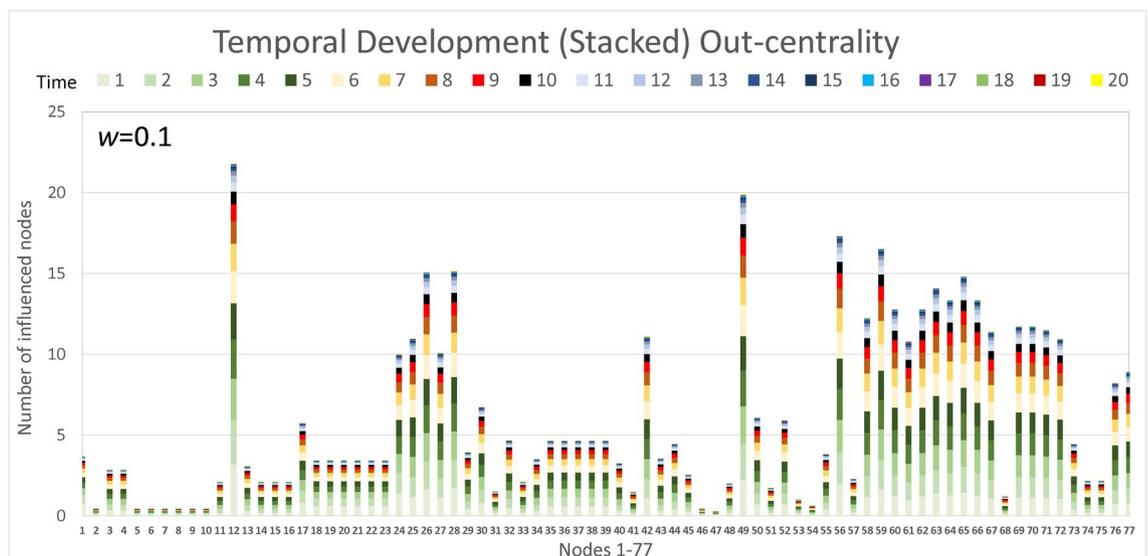


Fig. 8. Temporal development of out-centrality for low link weights. Temporal development of out-centrality illustrated as a stacked view for time values $T = 1, \dots, 20$ with link weights $w = 0.1$. The horizontal axis shows the nodes in the Les Misérables network of Fig. 2a and the vertical axis shows the expected number of influenced nodes calculated from Eq. 6. Path lengths $L \leq L_{max}$, where $L_{max} = 20$ are used in the calculations.

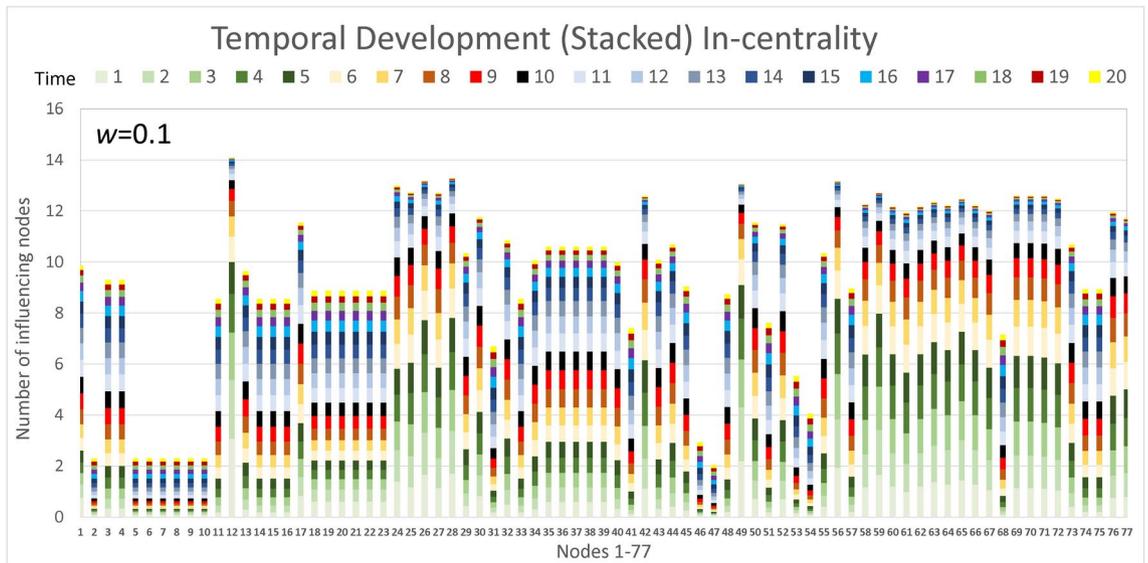


Fig. 9. Temporal development of in-centrality for low link weights. Temporal development of in-centrality illustrated as a stacked view for time values $T = 1, \dots, 20$ with link weights $w = 0.1$. The horizontal axis shows the nodes in the Les Misérables network of Fig. 2a and the vertical axis shows the expected number of influencing nodes calculated from Eq. 7. Path lengths $L \leq L_{max}$, where $L_{max} = 20$ are used in the calculations.

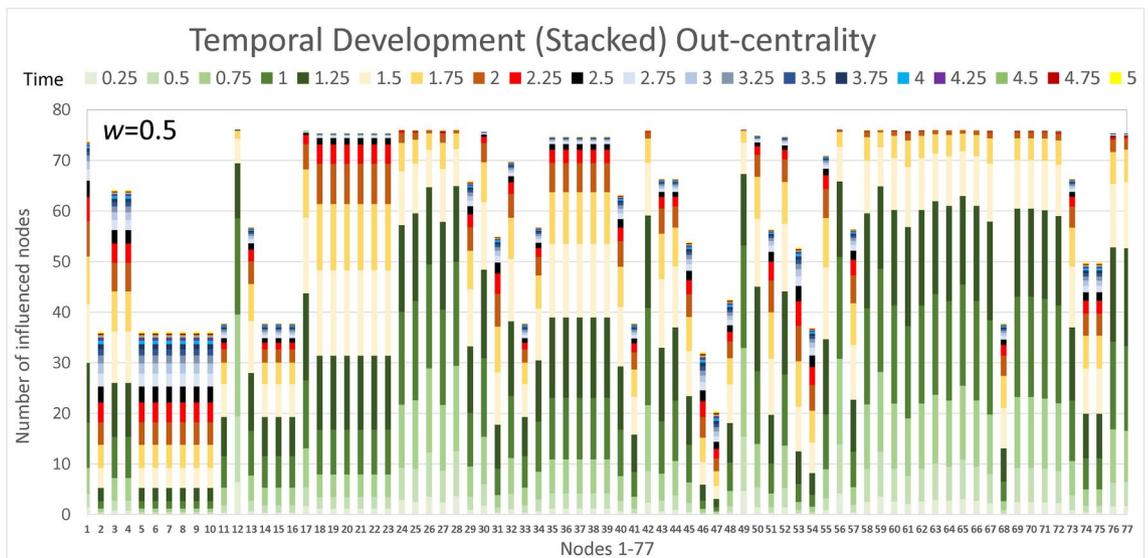


Fig. 10. Temporal development of out-centrality for high link weights. Temporal development of out-centrality illustrated as a stacked view for time values $T = 0.25, \dots, 5$ with link weights $w = 0.5$. The horizontal axis shows the nodes in the Les Misérables network of Fig. 2a and the vertical axis shows the expected number of influenced nodes calculated from Eq. 6. Path lengths $L \leq L_{max}$, where $L_{max} = 20$ are used in the calculations.

same similarity between the in-centrality values. This can be explained by the general properties of spreading processes in a network structure.

Effects depending on path length

The distribution that describes the dependency on the path length can be defined by a two-dimensional matrix which varies with time and path length. Time is a continuous variable, while path length is a discrete variable. In practice, we perform calculations in discrete time points. As an example of the use of the matrix, an additional factor may be included to account for decreasing influence as a function of path length. This would be applicable in both social influence and epidemic-spreading scenarios. Our model is global, with practical path lengths ranging between 1 to 20. To study local effects, low parameter values for the maximum path length can be used.

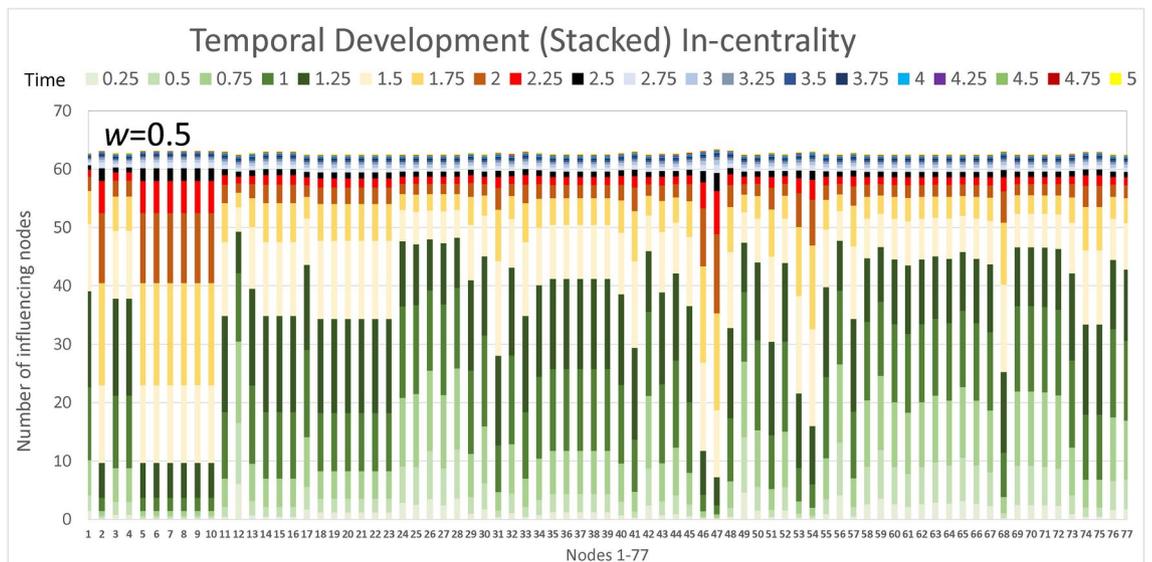


Fig. 11. Temporal development of in-centrality for high link weights. Temporal development of in-centrality illustrated as a stacked view for time values $T = 0.25, \dots, 5$ with link weights $w = 0.5$. The horizontal axis shows the nodes in the Les Misérables network of Fig. 2a and the vertical axis shows the expected number of influencing nodes calculated from Eq. 7. Path lengths $L \leq L_{max}$, where $L_{max} = 20$ are used in the calculations.

Here, we provide one numerical example to shed light on various possibilities of using different functions for describing effects that depend on the path length, in addition to the basic multiplicative rule in Eq. 3. The main dependency on the path length is through Eq. 3, which is used throughout in all of our examples. However, there is another possibility. We can generalise Eq. 2 for any kind of temporal distribution, including empirical distributions. We can even use other more general functions than probability distribution functions instead of Eq. 1.

Here, we provide a technical example to describe effects that depend on the path length. First, we define a $L_{max} \times T_{max}$ matrix $\mathcal{P}(L, T)$, $L = 1, \dots, L_{max}$, $T = 1, \dots, T_{max}$ where L_{max} and T_{max} denote the maximum values of path length and time. In our example we set $\mathcal{P}(L, T) = 1$ when $L = T$. We set $\mathcal{P}(L, T) = 0.75$ when $L = T - 1 = 1$ or $L - 1 = T = 1$ ($T > 1$ and $L > 1$). We set $\mathcal{P}(L, T) = 0.2$ when $L > T + 1$ and $\mathcal{P}(L, T) = 0.01$ when $T > L + 1$. This matrix is not a survival function of any probabilistic distribution because it is not an increasing function of T for a constant L . Numerical values of Table $\mathcal{P}(L, T)$ used in Figs. 12 and 13 are provided in Supplementary Table S2 online.

Figures 12 and 13 show the number of influenced nodes in the Les Misérables network of Fig. 2a for all nodes in the network as source nodes. Link weights are set $w = 0.1$ in Fig. 12 and $w = 0.5$ in Fig. 13 for all links in the network. The results in Fig. 12 for $w = 0.1$ have similarities with out-centrality results in Figs. 3 and 8. Similarly, the results in Fig. 13 have similarities with the results in Figs. 4 and 10.

The three curves show the results for time values $T = 1, 2$, and 3. The results for $T > 3$ remain almost the same. In Fig. 13, the most striking feature of the results is that $T = 3$ values are lower than $T = 2$ and for several source nodes $T = 3$ values are lower than $T = 1$. In general, it can be said that the shapes of the curves are roughly similar to previous results calculated with probability distributions in the sub-section Temporal Distributions. Because the temporal factors in Eq. 1 do not necessarily approach one when time approaches infinity $T \rightarrow \infty$, the values of matrix C introduced in the Model Section and derived results based on the matrix can depend on the functional form of $\mathcal{P}(L, T)$.

Conclusions

Our research presents a model for influence spreading on complex networks and simultaneously various metrics of these networks that can be utilised in different applications, such as social networks and outbreaks of epidemics. Our goal here is to improve the identification of important nodes and communities in the networks for various spreading processes. To accomplish this, we break down the analysis into two parts. Firstly, we determine the strength of influence between all directed node pairs in the network topology, which results in an influence-spreading matrix, or probability matrix. Secondly, we utilise the influence-spreading matrix to define centrality measures and discover community structures.

Our approach allows for the calculation of spreading probabilities between individual nodes or between any network structures. Measures of centrality and betweenness are defined consistently with other quantities in the model. Detailed-level modelling of the network structure and spreading process enables incorporating any form of temporal distribution in the spreading process without changing other parts of the model.

Our model is capable of analysing complex contagion processes that involve interactions between nodes in a network structure, including circular effects. In our earlier work, we have also explored a scenario where nodes are unable to form loops during the spreading process, which we classify as a simple contagion model.

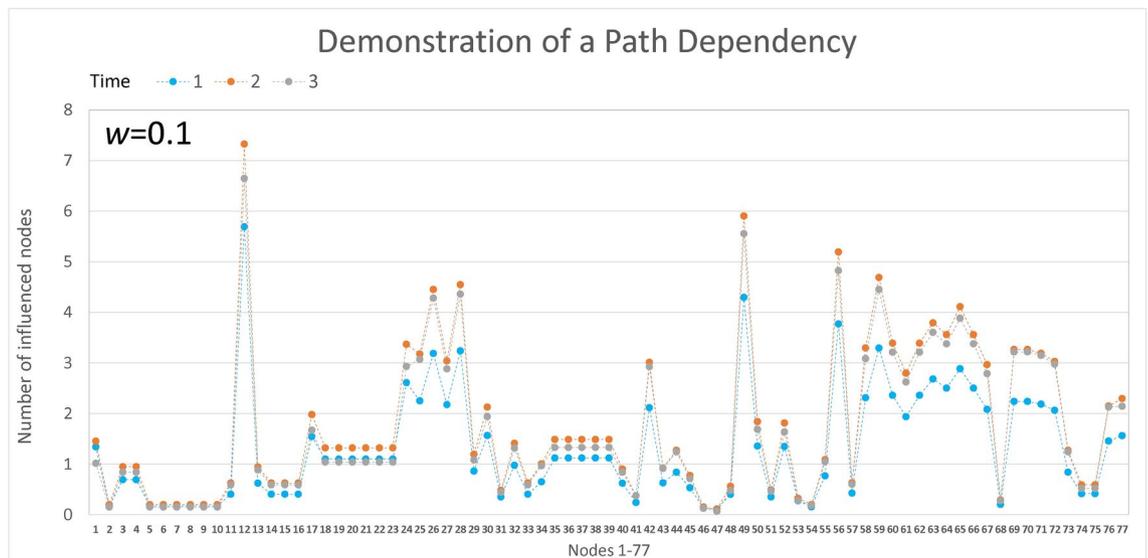


Fig. 12. A demonstration of temporal development of out-centrality for low link weights. Spreading probabilities as a function of the path length and time are provided in Supplementary Table S2 online. Temporal spreading illustrated for time values $T = 1, 2, 3$ with link weights $w = 0.1$. The horizontal axis shows the nodes in the Les Misérables network of Fig. 2a and the vertical axis shows the expected number of influenced nodes calculated from Eq. 6.

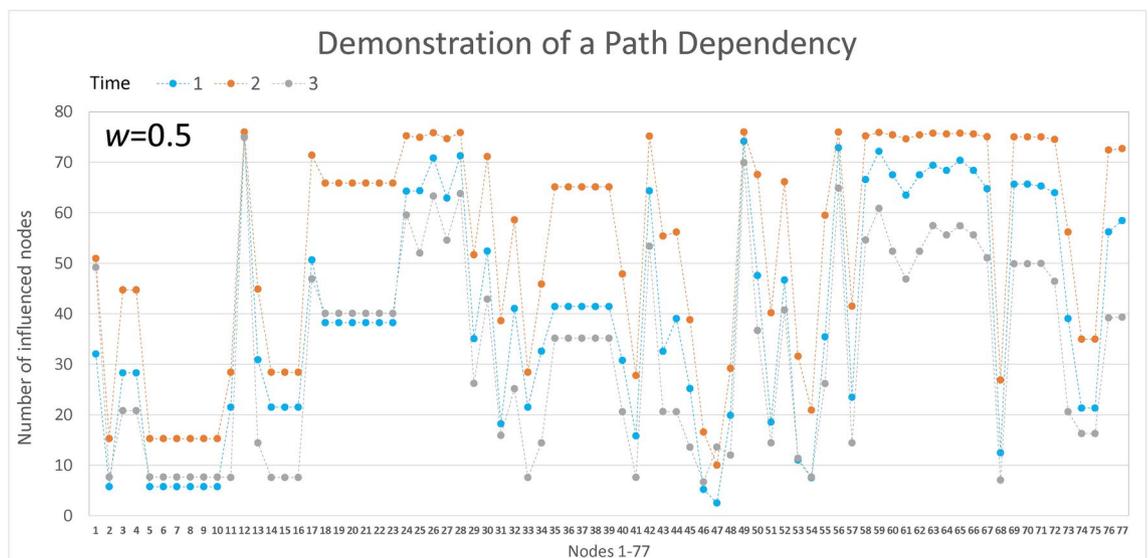


Fig. 13. A demonstration of temporal development of out-centrality for high link weights. Spreading probabilities as a function of the path length and time are provided in Supplementary Table S2 online. Temporal spreading illustrated for time values $T = 1, 2, 3$ with link weights $w = 0.5$. The horizontal axis shows the nodes in the Les Misérables network of Fig. 2a and the vertical axis shows the expected number of influenced nodes calculated from Eq. 6.

This model is useful for describing the spread of information and assessing the resilience of communication networks. In addition, we have unified the complex contagion and simple contagion models by incorporating breakthrough effects for all the nodes in the network¹⁵. The two models are limiting cases of each other when the breakthrough probability has extreme values of one or zero.

The proposed model has limitations in modelling time and state dependency in spreading processes. For instance, the model assumes similar breakthrough probabilities for all nodes in the network. In addition, the temporal distribution of spreading is a function of path length rather than being based on individual links. New features at the node and link levels can be incorporated into the algorithm using the same approach as the current version for utilising individual node and link weights. When studying epidemic outbreaks, the model

only accounts for two states for individual nodes in the network: susceptible and infected. To account for varying individual immunity during the incubation and infection periods, the model needs further development to include multi-variable modelling of time and state dependencies at the node level.

Our model for influence spreading and its applications involves defining centrality measures and quality functions for community detection. This is based on a detailed-level modelling of influence spreading on complex networks. The model utilises probabilities in its variables to ensure consistent results across different applications. The detailed-level model design is a novel method compared to other research. This method can remove inconsistencies and weaknesses reported in the literature concerning centrality measures based on shortest paths between nodes¹⁰ or inadequate modelling of network processes¹⁸. Sharing methods and metrics across the analysis can lead to new insights and understanding. The detailed-level modelling of influence spreading on complex networks can be applied to various areas of network analysis.

Data availability

The datasets used for the Les Misérables character network¹⁴ can be found in The Network Data Repository with Interactive Graph Analytics and Visualization⁶⁶, <https://networkrepository.com>.

Received: 31 January 2024; Accepted: 6 November 2024

Published online: 14 November 2024

References

- Newman, M. E. J. *Networks: An introduction* (Oxford University Press, Oxford, New York, 2018).
- Barabási, A.-L. Network science. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **371**, 20120375 (2013).
- Engel, U., Quan-Haase, A., Liu, S. X. & Lyberg, L. *Handbook of Computational Social Science* 1st edn. (Taylor & Francis, Routledge, 2021).
- Flache, A. et al. Models of social influence: Towards the next frontiers. *J. Artif. Soc. Soc. Simul.* <https://doi.org/10.18564/jasss.3521> (2017).
- Kim, S., Breen, J., Dudkina, E., Poloni, F. & Crisostomi, E. On the effectiveness of random walks for modeling epidemics on networks. *PLoS ONE* **18**, e0280277. <https://doi.org/10.1371/journal.pone.0280277> (2023).
- Kuikka, V., Aalto, H., Ijäs, M. & Kaski, K. K. Efficiency of algorithms for computing influence and information spreading on social networks. *Algorithms[SPACE]* <https://doi.org/10.3390/a15080262> (2022).
- Kuikka, V. Influence spreading model used to analyse social networks and detect sub-communities. *Comput. Soc. Netw.* **5**, 12–15. <https://doi.org/10.1186/s40649-018-0060-z> (2018).
- Kuikka, V. Modelling community structure and temporal spreading on complex networks. *Comput. Soc. Netw.* **8**, 13. <https://doi.org/10.1186/s40649-021-00094-z> (2021).
- Centola, D. & Macy, M. Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**, 702–734. <https://doi.org/10.1086/521848> (2007).
- Landherr, A., Friedl, B. & Heidemann, J. A critical review of centrality measures in social networks. *Bus. Inf. Syst. Eng.* **2**, 371–385. <https://doi.org/10.1007/s12599-010-0127-3> (2010).
- Gómez, S. Centrality in networks: Finding the most important nodes. *Bus. Consum. Anal.: New Ideas[SPACE]* https://doi.org/10.1007/978-3-030-06222-4_8 (2019).
- Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113. <https://doi.org/10.1103/PhysRevE.69.026113> (2004).
- Fortunato, S. & Hric, D. Community detection in networks: A user guide. *Phys. Rep.* **659**, 1–44. <https://doi.org/10.1016/j.physrep.2016.09.002> (2016).
- Riolo, M. A. & Newman, M. Consistency of community structure in complex networks. *Phys. Rev. E* **101**, 052306. <https://doi.org/10.1103/PhysRevE.101.052306> (2020).
- Almiala, I., Aalto, H. & Kuikka, V. Influence spreading model for partial breakthrough effects on complex networks. *Phys. A* **630**, 129244. <https://doi.org/10.1016/j.physa.2023.129244> (2023).
- Kuikka, V. Modelling epidemic spreading in structured organisations. *Phys. A* **592**, 126875. <https://doi.org/10.1016/j.physa.2022.126875> (2022).
- Kuikka, V. Detecting overlapping communities based on influence-spreading matrix and local maxima of a quality function. *Computation* **12**, 85. <https://doi.org/10.3390/computation12040085> (2024).
- Bockholt, M. & Zweig, K. A. Why we need a process-driven network analysis. In *Complex Networks and Their Applications VIII* (eds Cherifi, H. et al.) 81–93 (Springer International Publishing, Cham, 2020). https://doi.org/10.1007/978-3-030-36683-4_7.
- Borgatti, S. P. Centrality and network flow. *Soc. Netw.* **27**, 55–71. <https://doi.org/10.1016/j.socnet.2004.11.008> (2005).
- Horsevad, N., Mateo, D., Kooij, R., Barrat, A. & Bouffanais, R. Transition from simple to complex contagion in collective decision-making. *Nat. Commun.* **13**, 1442. <https://doi.org/10.1038/s41467-022-28958-6> (2022).
- Kirkley, A., Cantwell, G. T. & Newman, M. E. J. Belief propagation for networks with loops. *Sci. Adv.* **7**, eabf1211. <https://doi.org/10.1126/sciadv.abf1211> (2021).
- Liu, C., Zhan, X.-X., Zhang, Z.-K., Sun, G.-Q. & Hui, P. M. How events determine spreading patterns: Information transmission via internal and external influences on social networks. *New J. Phys.* **17**, 113045 (2015).
- Bonifazi, G. et al. Investigating community evolutions in tiktok dangerous and non-dangerous challenges. *J. Inf. Sci.[SPACE]* <https://doi.org/10.1177/01655515221116519> (2022).
- Nowzari, C., Preciado, V. M. & Pappas, G. J. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Syst. Mag.* **36**, 26–46. <https://doi.org/10.1109/MCS.2015.2495000> (2016).
- de Arruda, G. F., Rodrigues, F. A. & Moreno, Y. Fundamentals of spreading processes in single and multilayer complex networks. *Phys. Rep.* **756**, 1–59. <https://doi.org/10.1016/j.physrep.2018.06.007> (2018).
- Newman, M. Message passing methods on complex networks. *Proc. R. Soc. A* **479**, 20220774. <https://doi.org/10.1098/rspa.2022.0774> (2023).
- Zhang, Y., Lucas, M. & Battiston, F. Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes. *Nat. Commun.* **14**, 1605. <https://doi.org/10.1038/s41467-023-37190-9> (2023).
- Iacopini, I., Petri, G., Barrat, A. & Latora, V. Simplicial models of social contagion. *Nat. Commun.[SPACE]* <https://doi.org/10.1038/s41467-019-10431-6> (2019).
- Centola, D. *How Behavior Spreads: The Science of Complex Contagions* (Princeton University Press, 2018).
- Min, B. & Miguel, M. Competing contagion processes: Complex contagion triggered by simple contagion. *Sci. Rep.[SPACE]* <https://doi.org/10.1038/s41598-018-28615-3> (2018).

31. Guilbeault, D., Becker, J. & Centola, D. Complex contagions: A decade in review. *Complex spreading phenomena in social systems: Influence and contagion in real-world social networks*, https://doi.org/10.1007/978-3-319-77332-2_1 (2018).
32. Khajanchi, S., Sarkar, K., Mondal, J., Nisar, K. S. & Abdelwahab, S. F. Mathematical modeling of the covid-19 pandemic with intervention strategies. *Res. Phys.* **25**, 104285. <https://doi.org/10.1016/j.rinp.2021.104285> (2021).
33. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979. <https://doi.org/10.1103/RevModPhys.87.925> (2015).
34. Fernandez Peralta, A., Kertész, J. & Iñiguez, G. Opinion dynamics in social networks: From models to data. *Handbook of Computational Social Science* (2023).
35. Perra, N. Non-pharmaceutical interventions during the covid-19 pandemic: A review. *Phys. Rep.* **913**, 1–52. <https://doi.org/10.1016/j.physrep.2021.02.001> (2021).
36. Devia, C. A. & Giordano, G. Probabilistic analysis of agent-based opinion formation models. *Sci. Rep.* **13**, 20152. <https://doi.org/10.1038/s41598-023-46789-3> (2023).
37. Devia, C. A. & Giordano, G. A framework to analyze opinion formation models. *Sci. Rep.* **12**, 13441. <https://doi.org/10.1038/s41598-022-17348-z> (2022).
38. Bick, C., Goodfellow, M., Laing, C. R. & Martens, E. A. Understanding the dynamics of biological and neural oscillator networks through exact mean-field reductions: a review. *J. Math. Neurosci.* **10**, 9. <https://doi.org/10.1186/s13408-020-00086-9> (2020).
39. Nekovee, M., Moreno, Y., Bianconi, G. & Marsili, M. Theory of rumour spreading in complex social networks. *Phys. A* **374**, 457–470. <https://doi.org/10.1016/j.physa.2006.07.017> (2007).
40. Lambiotte, R., Rosvall, M. & Scholtes, I. From networks to optimal higher-order models of complex systems. *Nat. Phys.* **15**, 313–320. <https://doi.org/10.1038/s41567-019-0459-y> (2019).
41. Kozitsin, I. V. A general framework to link theory and empirics in opinion formation models. *Sci. Rep.* **12**, 5543. <https://doi.org/10.1038/s41598-022-09468-3> (2022).
42. Helfmann, L., Djurdjevac Conrad, N., Lorenz-Spreen, P. & Schütte, C. Modelling opinion dynamics under the impact of influencer and media strategies. *Sci. Rep.* **13**, 19375. <https://doi.org/10.1038/s41598-023-46187-9> (2023).
43. Bogomolnaia, A. & Jackson, M. O. The stability of hedonic coalition structures. *Games Econ. Behav.* **38**, 201–230. <https://doi.org/10.1006/game.2001.0877> (2002).
44. Rosenthal, R. W. A class of games possessing pure-strategy nash equilibria. *Int. J. Game Theory* **2**, 65–67 (1973).
45. Le Breton, M., Shapoval, A. & Weber, S. A game-theoretical model of the landscape theory. *J. Math. Econ.* **92**, 41–46. <https://doi.org/10.1016/j.jmateco.2020.11.004> (2021).
46. Le Breton, M. & Weber, S. Games of social interactions with local and global externalities. *Econ. Lett.* **111**, 88–90. <https://doi.org/10.1016/j.econlet.2011.01.012> (2011).
47. Axelrod, R. & Bennett, D. S. A landscape theory of aggregation. *Br. J. Political Sci.* **23**, 211–233. <https://doi.org/10.1017/S000712340000973X> (1993).
48. Freeman, L. C., Borgatti, S. P. & White, D. R. Centrality in valued graphs: A measure of betweenness based on network flow. *Soc. Netw.* **13**, 141–154. [https://doi.org/10.1016/0378-8733\(91\)90017-N](https://doi.org/10.1016/0378-8733(91)90017-N) (1991).
49. Bonifazi, G. et al. An approach to detect backbones of information diffusers among different communities of a social platform. *Data Knowl. Eng.* **140**, 102048. <https://doi.org/10.1016/j.datak.2022.102048> (2022).
50. Friedkin, N. E. Theoretical foundations for centrality measures. *Am. J. Sociol.* **96**, 1478–1504. <https://doi.org/10.1086/229694> (1991).
51. Fink, C. G. et al. A centrality measure for quantifying spread on weighted, directed networks. *Available at SSRN[SPACE]* <https://doi.org/10.2139/ssrn.4415860> (2023).
52. Ullah, A., Sheng, J., Wang, B., Din, S. U. & Khan, N. Leveraging neighborhood and path information for influential spreaders recognition in complex networks. *J. Intell. Inf. Syst.* [SPACE] <https://doi.org/10.1007/s10844-023-00822-z> (2023).
53. Bockholt, M. & Zweig, K. A. A systematic evaluation of assumptions in centrality measures by empirical flow data. *Soc. Netw. Anal. Min.* **11**, 1–30. <https://doi.org/10.1007/s13278-021-00725-3> (2021).
54. Bockholt, M. & Zweig, K. A. Process-driven betweenness centrality measures. *Netw. Intell. Meets User Cent. Soc. Med. Netw.* **4**, 17–33. https://doi.org/10.1007/978-3-319-90312-5_2 (2018).
55. Rajeh, S. & Cherifi, H. Ranking influential nodes in complex networks with community structure. *PLoS ONE* **17**, e0273610. <https://doi.org/10.1371/journal.pone.0273610> (2022).
56. Rajeh, S., Savonnet, M., Leclercq, E. & Cherifi, H. How correlated are community-aware and classical centrality measures in complex networks? In *International Conference on Complex Networks*, 120–132. https://doi.org/10.1007/978-3-030-81854-8_11 (Springer, 2021).
57. Chen, Y., Zhao, P., Li, P., Zhang, K. & Zhang, J. Finding communities by their centers. *Sci. Rep.* **6**, 24017. <https://doi.org/10.1038/srep24017> (2016).
58. Lee, C. & Wilkinson, D. J. A review of stochastic block models and extensions for graph clustering. *Appl. Netw. Sci.* **4**, 1–50. <https://doi.org/10.1007/s41109-019-0232-2> (2019).
59. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An open source software for exploring and manipulating networks. <https://doi.org/10.1609/icwsm.v3i1.13937> (2009).
60. Jerdee, M., Kirkley, A. & Newman, M. Normalized mutual information is a biased measure for classification and community detection. *arXiv e-prints* arXiv–2307. <https://doi.org/10.48550/arXiv.2307.01282> (2023).
61. Leão, J. C., Laender, A. H. & de Melo, P. O. V. Overcoming bias in community detection evaluation. *J. Inf. Data Manag.* [SPACE] <https://doi.org/10.5753/jidm.2020.2018> (2020).
62. Fortunato, S. & Barthelemy, M. Resolution limit in community detection. *Proc. Natl. Acad. Sci.* **104**, 36–41. <https://doi.org/10.1073/pnas.0605965104> (2007).
63. Ghasemiefteh, G., Ebrahimi, R. & Gao, J. Complex contagion and the weakness of long ties in social networks: Revisited. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce, EC'13*, 507–524. <https://doi.org/10.1145/2482540.2482550> (Association for Computing Machinery, New York, NY, USA, 2018).
64. Almiala, I. & Kuikka, V. Similarity of epidemic spreading and information network connectivity mechanisms demonstrated by analysis of two probabilistic models. *AIMS Biophys.* **10**, 173–183. <https://doi.org/10.3934/biophys.2023011> (2023).
65. Kuikka, V. & Rantanen, H. Resilience of multi-layer communication networks. *Sensors* **23**, 86. <https://doi.org/10.3390/s23010086> (2022).
66. Rossi, R. A. & Ahmed, N. K. The network data repository with interactive graph analytics and visualization. In *AAAI* (2015).
67. Kuikka, V. Influence spreading model used to community detection in social networks. In Cherifi, C., Cherifi, H., Karsai, M. & Musolesi, M. (eds.) *Complex Networks & Their Applications VI, Proceedings of Complex Networks 2017*, 202–215. https://doi.org/10.1007/978-3-319-72150-7_17 (Springer International Publishing, Cham, 2018).
68. Ijäs, M., Levijoki, J. & Kuikka, V. Scalable algorithm for computing influence spreading probabilities in social networks. In *5th European Conference on Social Media (ECSM 2018)* (2018).

Acknowledgements

We acknowledge that the efficient software implementation of the Complex Contagion Algorithm⁶ is based on the study by Vesa Kuikka⁶⁷ and result of development in cooperation with Janne Levijoki and Matias Ijäs⁶⁸.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-79182-9>.

Correspondence and requests for materials should be addressed to V.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024