Modelling community structure and temporal spreading on complex networks

Vesa Kuikka (vesa.kuikka@mil.fi)
Finnish Defence Research Agency, Tykkikentäntie 1, PO BOX 10, 11311 Riihimäki, Finland

Research

Keywords: Complex networks, Community detection, Influence spreading model, Network connectivity, Probability matrix, E-mail forwarding process

Posted Date: March 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-16363/v1

Version of Record: A version of this preprint was published on March 18th, 2021. See the published version at https://doi.org/10.1186/s40649-021-00094-z.

Modelling community structure and temporal spreading on complex networks

Vesa Kuikka[0000-0002-3677-816X]

[1] Finnish Defence Research Agency, Tykkikentäntie 1,

PO BOX 10, 11311 Riihimäki, Finland

vesa.kuikka@mil.fi

Abstract. We present methods for analysing hierarchical and overlapping community structure and spreading phenomena on complex networks. Different models can be developed for describing static connectivity or dynamical processes on a network topology. In this study, classical network connectivity and influence spreading models are used as examples for network models. Analysis of results is based on a probability matrix describing interactions between all pairs of nodes in the network. One popular research area has been detecting communities and their structure in complex networks. The community detection method of this study is based on optimising a quality function calculated from the probability matrix. The same method is proposed for detecting underlying groups of nodes that are building blocks of different sub-communities in network structure. We present different quantitative measures for comparing and ranking solutions of the community detection algorithm. These measures describe properties of sub-communities: strength of a community, probability of formation and robustness of composition. We illustrate the community detection methods with two small network topologies. In case of network spreading models, time development of spreading in the network can be studied. Two different temporal spreading distributions demonstrate the methods with three real-world social networks of different sizes. The Poisson distribution describes a random response time and the e-mail forwarding distribution describes a process of receiving and forwarding messages.

Keywords: Complex networks, Community detection, Influence spreading model, Network connectivity, Probability matrix, E-mail forwarding process.

1 Introduction

Network science studies a wide range of properties and phenomena on complex networks [1]. Applications are on social networks, biomedical networks, technical networks and many other fields. Some practical examples of research

questions are studying more realistic community detection methods, investigating processes on networks, and defining different quantitative measures for comparing and ranking network structures and processes. In this paper we propose one methodology for studying both community structure and temporal spreading on complex networks.

In many complex systems, dependencies can be modelled with networks consisting of nodes and links [2]. Nodes can have different roles as central influencers or mediators depending on their location in the network structure. Links and nodes can have attributes that describe their static or dynamical properties. These networks are complex because nodes typically are connected in a complex way in network topology and complex phenomena like phase transitions and different sub-structures can emerge [3]. Community detection methods have been developed in order to understand formation of groups in social networks or for categorising sub-systems in technological and biological networks. Community detection methods can be based on spreading models or static network models.

Brief general introductions to community detection methods and processes on networks are provided in Sections 'Community detection methods' and 'Processes on complex networks'. Standard centrality and betweenness measures are introduced in Section 'Centrality measures' together with some comments on some less commonly used centrality measures related to the definitions of centrality proposed in this paper. Two example network models, an influence spreading model and the classical network connectivity model, are introduced in Section 'Example network models'. Rest of the paper has two main Sections 'Structure in complex networks' and 'Modelling temporal spreading on networks'. Section 'Structure in complex networks' is divided into three sub-sections 'Community detection method', 'Influence spreading model', and results in 'Detected communities and their structure'. Results include discussion about sensitivity to model parameter values, quantitative measures for ranking communities according to their strength, probability of formation and robustness of composition. A new method is proposed for discovering underlying groups of nodes (building blocks) that are usually found together in community structures. The main conclusions are presented in the last section.

1.1 Community detection methods

Community detection is one of the most important applications of modelling complex networks. A great number of different community detection methods and algorithms have been published in the literature [4]. One problem with various approaches is the lack of a commonly accepted definition of a community. On the other hand, one definition for all possible purposes may not be possible because of different empirical data available and requirements in

applications. A more realistic approach would be first to categorise and define more general concepts of complex networks, such as types of networks, processes, function rules and interactions. In many cases, the concept of a community is not defined exactly, instead the used method and algorithm define the concept implicitly. In this way, the number of definitions of a community will increase along with the number of methods and new versions of algorithms. Mathematical methods and algorithms for detecting communities in complex network topologies have been reviewed and presented in [4, 5, 6, 7].

Modularity maximisation, classical graph partitioning, spectral graph partitioning and several information theoretic methods are examples in the wide context of community detection methods [1, 2]. The classical graph partitioning is the problem of dividing the nodes of a network into a given number of non-overlapping groups of given sizes such that the number of links between groups is minimised. Modularity measures the strength of division of a network into modules. One definition of a community is a locally dense connected sub-graph in a network [1]. Modularity has been defined as the fraction of links falling within the given groups minus the expected fraction if links were distributed at random. In order to compute the numerical value of modularity, each link is cut into two halves, called stubs. The expected number of links is computed by rewiring stubs randomly with any other stub in the network, except itself, but allowing self-loops when a stub is rewired to another stub from the same node. Mathematically, modularity can be expressed as

$$M = \frac{1}{2m}\sum_{vw}\left(A_{vw} - \frac{k_v k_w}{2m}\right)\frac{s_v s_w + 1}{2}. \quad (1)$$

In Eq. (1) v and w are nodes in the network, $2m$ is the number of stubs in the network, k_v is the node degree of node v, $A_{vw} = 1$ means that there is a link between nodes v and w, and $A_{vw} = 0$ means that there is no link between the two nodes. Matrix A is called the adjacency matrix. Membership variable s_v indicates if node v belongs to a community: $s_v = 1$ if node v belongs to community 1, and $s_v = -1$ if node v belongs to community 2. [2]

Equation (1) holds for partitioning into two modules but it can be generalised for partitioning into a desired number of modules. In matrix terms Eq. (1) is

$$M = \frac{1}{4m}s^T B s,$$

where $B_{vw} = A_{vw} - \frac{k_v k_w}{2m}$ is called the modularity matrix. The equation for M is similar in form to an expression used in spectral partitioning of graphs for the cut size of a network in terms of the graph Laplacian. This similarity can be used for deriving a spectral algorithm for community detection. The eigenvector corresponding to the largest eigenvalue of the modularity matrix assigns nodes to communities according to the signs of the vector elements. [2]

The Louvain algorithm [8] and Infomap [9] are two fast algorithms for community detection that have been briefly described in [1]. These algorithms have gained popularity because of their suitability for identifying communities in very large networks. Both algorithms optimise a quality function. For the Louvain algorithm the quality function is modularity and for Infomap an entropy-based measure. In the Louvain algorithm modularity is optimised by local changes of the modularity measure and communities are obtained by aggregating the modules to build larger communities. Infomap compresses the information about a random walker exploring the graph. [1]

Stochastic blockmodels have been used as a method for detecting community structure in networks and also for generating synthetic benchmark networks [10]. Many community detection methods discover also hierarchical and overlapping sub-communities in complex networks [11].

1.2 Processes on complex networks

In a dynamical system, state of the system changes over time according to some given rules [3]. Dynamical processes on complex networks has growing interest because of its wide scope of applications. Epidemic spreading in populations [12], influence spreading in social networks [13] and the flow of traffic on roads are important practical applications [3]. Standard approaches of studying dynamical processes on networks rely on simulations because analytical mathematical expressions are not available or they are very complicated. However, remarkable research has been conducted in percolation theory where analytical solutions exist for some network topologies. These methods are not directly applicable for modelling detailed topology of empirical networks. [2]

By definition, processes on networks have a time dependency. Two examples of processes on networks are spreading on network structure and changing network topology [14]. Spreading processes start from a node or a set of nodes in the network and propagate between nodes via links in the network structure. Also, the network topology may change during the spreading process. Changes in network structure and changes in link and node attribute values are common in many applications. For example, virus spreading in computer networks or in human social networks, are slowed down with virus protection software or vaccination programs in human population.

104 Spreading processes on networks cover a variety of situations because processes can depend on states of other
105 nodes or links in the network. For example, virus spreading may not be possible or it is only partial in case nodes are
106 immunised as a result of a previous contamination [12]. Another example is when information or rumours are spread
107 more actively when heard first time compared to later versions of the same information. In social systems many
108 overlapping processes are simultaneously influencing our beliefs and opinions [15].
109

110 **1.3 Centrality measures**

111 Nodes in a network can have different roles as central influencers, mediators or peripheral nodes. Degree is the
112 simplest closeness centrality measure. Degree centrality of a node is defined as the number of nodes connected to it.
113 It is a local measure and does not take into account node's position in the network. Closeness centrality of a node
114 measures how central or influential the node is in respect to other nodes. Betweenness centrality measures the role of
115 a node as a proxy between other nodes in the network. Closeness and betweenness centrality measures have many
116 variants and they depend on the research question of a particular application. Mathematical theories provide more
117 centrality measures such as the eigenvector centrality. In-centrality and out-centrality can be defined both for directed
118 and undirected networks. Normally, betweenness of a node is not defined for inward and outward directions. Review
119 article [16] presents definitions and descriptions of the most commonly used eight centrality measures. Many other
120 centrality measures and their variants have been proposed in the literature.
121 Standard closeness centrality is based on the inverse sum of the shortest distances to the other nodes of the network.
122 Distance in the network is defined as the minimum number of links needed to move from one node to another.

123 $$C_i^{(clos1)} = \left(\sum_{j=1}^{N} d_{i,j} \right)^{-1}.$$

124 A normalised version can be obtained by multiplying the formula by $N-1$ where N is the number of nodes in the
125 network. The factor $N-1$ is justified because the distance from a node to itself is zero.
126 If the network is not connected, some of the distances are infinite and the closeness centrality of all nodes becomes
127 zero. An alternative definition avoids this by defining another version of the closeness centrality measure as

128 $$C_i^{(clos2)} = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} {d_{i,j}}^{-1}$$

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

$$C_i^{(betw)} = \frac{1}{(N-1)(N-2)} \sum_{\substack{s,t=1 \\ s \neq t \neq i}}^{N} \frac{\sigma_{s,t}(i)}{\sigma_{s,t}}.$$

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

In addition to the standard measures, many other definitions have been proposed in the literature. Some of them are related to the centrality and betweenness measures proposed in Section 'Influence spreading model' of this paper. One of the centrality measures is the Katz centrality measure [17]. The Katz centrality measure generalises the degree centrality and the closeness centrality by taking into account not only the immediate neighbours or not only the shortest paths from a node to other nodes. The Katz centrality is defined as

$$C_i^{(Katz)} = \sum_{k=1}^{\infty} \sum_{j=1}^{N} \alpha^k (A^k)_{ji}.$$

The power of adjacency matrix A accounts for the number of paths of length k between every pair of nodes. A decay parameter $\alpha < 1$ is introduced to weight the contributions of nodes at increasing path lengths.

Eigenvector centrality is a measure which depends recursively on the centralities of node's neighbours. Katz centrality, hubs and authorities centrality and PageRank are variants of eigenvector centrality. The basic eigenvector centrality measure is used only for undirected networks whereas the three variants are also appropriate for directed networks. Hubs and authorities centrality assigns to each node two different measures both for sending and receiving influence. For undirected networks this definition coincides with the basic eigenvector centrality and the distinction between hubs and authorities disappears. [16]

1.4 Example network models

We use two network models to demonstrate the community detection method of Section 'Community detection method': the classical network connectivity model [18] and the influence spreading model proposed in [19, 20, 21]. The influence spreading model and the network connectivity model have been designed for different application areas. The primary use of the influence spreading model is for behaviour and opinion spreading and the connectivity model is commonly used for modelling communication networks.

Temporal spreading is only possible, by definition, in network spreading models. This is the reason why we present a method for modelling temporal distributions for the influence spreading model and not for the network connectivity model. Next, we present briefly the main features of the two models. Here, we give an general idea of the influence spreading model. More detailed descriptions are in Section 'Influence spreading model' and in [16]. The network connectivity model is included in this paper mostly for comparison purposes and more details of the model can be found in text books or research articles [18].

1.4.1 Influence spreading model

Influence spreading models are designed for describing complex social interactions in social network structure [12, 25]. These interactions propagate via connections, or paths, between people. We assume that information content can change and ways of social influence are developing during the spreading process. We allow repeated attempts of influence from a source node to target nodes via all alternative paths. This includes also loops, but not self-loops, where one node can occur several times on a path. Social influence and spreading of beliefs can be modelled with these kind of processes. In the process of spreading information and news, loops are less probable.

Other features of the model are: weighted links and nodes, directed links, and the possibility of using different forms of temporal survival distributions as a function of the number of links between a source node and a target node. All the parameters have real-world interpretations. Link and node weights are interpreted as spreading probabilities of forwarding influence between neighbouring nodes and over a node, respectively. Spreading probabilities between all pairs of nodes in a network can be calculated from node and link weighting values and the temporal survival distribution function. [20]

The influence spreading model used in this paper has been presented previously in [19, 20, 21]. The model has analytical expressions for spreading probabilities via different paths from a source node to target nodes in the network

topology. In this version of the model, the rate of spreading is assumed to be independent of the state of the network and its elements. To avoid double counting effects, common paths at their beginning from one source node to a target node are taken into account by applying probability theory. In other respects, different paths are assumed to be independent. As a result, in case paths join or cross later, the spreading process is not affected.

We illustrate the model with an example of combining two paths of lengths L_1 and L_2 with a common path of length L_3. The probability of spreading via the two paths as a function of time T is

$$C_{s,t} = P_{L_1}(T) + P_{L_2}(T) - \frac{P_{L_1}(T)P_{L_2}(T)}{P_{L_3}(T)}.$$

The probability value of spreading via the two paths form source node s to target node t is denoted by $C_{s,t}$. Probabilities $P_{L_i}, i = 1,2,3$ contain the relevant link and node weighs on the paths and the numerical values of a temporal survival distribution function. Later, in our numerical calculations we use limiting values of time T approaching infinity. This describes equilibrium of a network as the final sate of spreading. [20]

An important property of a process is the possibility of loops. If recurrent visits on a node are allowed, one node can be visited several times during the process. In this paper, we assume that loops are allowed. In the algorithm, it is also possible to set the maximum number of visits $V = 1, \dots, L$, where L is of the maximum path length of computing.

1.4.2 Network connectivity model

The classical network connectivity model is designed for describing reliability of communication networks [18]. If the reliability values between all neighbouring pairs of nodes in the network are known, reliability values between any pairs of nodes in the network can be computed. Reliability is identified with the probability of an operational connection in a time unit. From the general reliability theory [18] the reliability of network V is

$$r(V) = \sum_{S \in O} \prod_{e \notin S} (1 - p_e) \prod_{e \in S} p_e,$$

where S is a set of links where the network is connected, and O is the set of all connected states of the network. Links are denoted by e and the probability of an operational link is denoted by p_e. If the probabilities p_e are equal,

$$r(V) = 1 - \prod_{h=1}^{N_L} \left(\sum_{s=0}^{h} (-1)^{h-s} \binom{N_L - s}{N_L - h} H_S \right) p_e^h ,$$

where H_s is the sum of indicator functions when the number of broken links is s. The above equations are polynomials of the order of the number of links N_L in the network. In this form, the equations describe reliability of the entire network. In our case we apply the results for pairs of nodes by only taking the relevant terms in the summations.

2 Structure in complex networks

In this section, we present methods for studying structure in complex networks with the help of network models and a community detection method. The primary network model in this study is the influence spreading model presented in [20]. The community detection model is also presented in [20]. Demonstrations of the model are for the equilibrium state with time approaching infinity and temporal spreading distributions are not presented in this section. Hierarchical and overlapping sub-communities are detected and they are illustrated as hierarchical graphs. We propose several quality measures for ranking the detected divisions and sub-communities. Calculations with different model parameters and link weights with random components support the assumption that justified conclusions can be made from the results despite the fact that link weight values are difficult to estimate for real-world social networks. In addition, the methodology can be used for discovering building blocks for communities and sub-communities. Building blocks are groups of nodes that are usually found together. The method of detecting building blocks is not restricted to the community detection method of Section 'Community detection method'.

2.1 Community detection method

In this paper, we take an approach, where instead of performing community detection based on the adjacency matrix of a graph, an influence matrix is first constructed that contains information about a social influence process over the paths of a graph [20]. An element C_{vw} of the influence matrix C accounts for interactions over all the paths from source node v to target node w. In order to study local interactions in the network, maximum path length can be set in the algorithm. In addition, we take into account that interactions in communities can mean different things depending on the processes that are supposed to operate on a network. This is demonstrated by substituting the social influence matrix with the network connectivity matrix known from the classical communication theory [18].

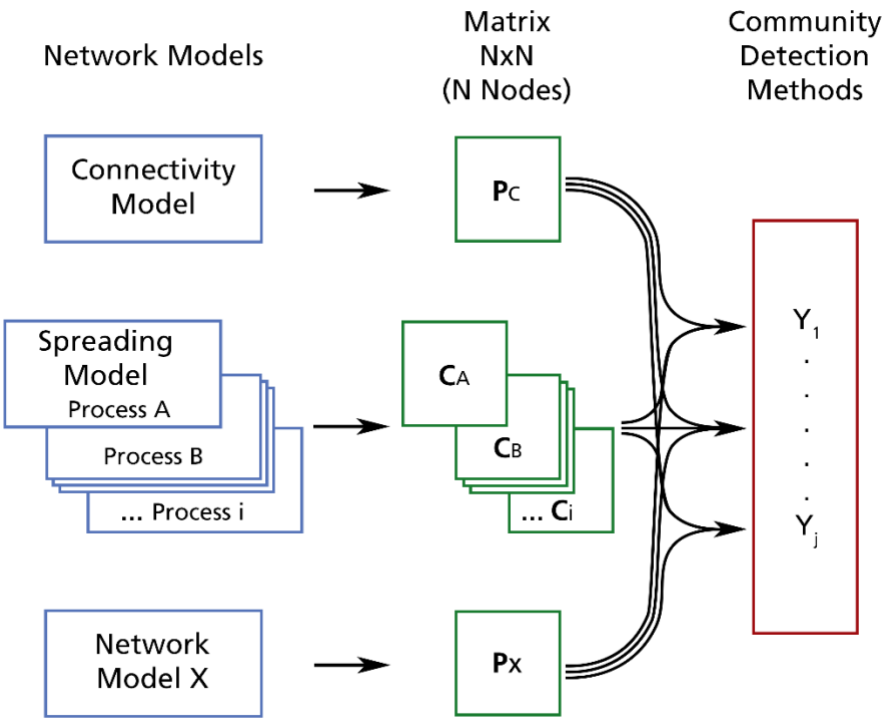


Fig. 1. General procedure of using different network models and community detection models. Information from network models are mediated with a probability matrix describing influence or connectivity between all node pairs in the network.

In the community detection algorithm a sum of rows and columns of matrix C is used as the quality function. Rows and columns are included in the sum that correspond to node pairs in the community, and node pairs not in the community. This measure is different from the modularity M of Eq. (1). Typically, a node has higher influence on neighbouring nodes compared to nodes that are far away in the network topology. Influence is also increasing with the number of alternative paths between nodes. Most community detection methods consider only the local influence among nodes in network structure. More accurate results can be obtained when longer path lengths are included in the model and calculations. In order to balance between the increasing number of alternative paths and distance between a source node and a target node, weighting factors are used to describe probabilities of influence via links between two neighbouring nodes in the network. In static connectivity models the probability of functioning connection plays the role of interaction. Weighting factors for links or nodes, or both, together with the network topology are the main

input data for network models. In dynamical models, spreading time distribution and time dependency of node and link attribute values describe influence spreading and changing network structure.

The community detection method separates the network model and the community detection model. This means that the same community detection algorithm can be used with different network models. We present two examples of network models: a network connectivity model [18] and an influence spreading model [20]. In both cases, we use the same community detection algorithm. In these two examples, interactions between nodes are described as spreading probabilities or probabilities of functioning connections. Technically, interactions between pairs of nodes in the network are expressed in a $N \times N$ –dimensional probability matrix C where N is the number of nodes in the network. Matrix C contains computed values for all pairs of nodes in the network, not just for neighbouring nodes.

The method detects many kinds of structures in topological complex networks: non-overlapping, overlapping and hierarchical community structure. As a special case, communities consisting of two or more distinct sub-communities that have no direct contact, can be discovered.

The method is based on searching local maxima of a community influence measure computed from the probability matrix elements $C_{s,t}, s, t = 1, \dots, N$. Our basic model has the following form for the community influence measure:

$$P = \sum_{s,t \in V} C_{s,t} + \sum_{s,t \in (1-V)} C_{s,t} \tag{2}$$

In Eq. (2) the first summation is over the pairs of nodes in a subset V of nodes in the network and the second summation is over the remaining partition $(1 - V)$ of nodes. Cross terms are ignored in this version of the model because they describe interactions between the two partitions of the division and are not directly involved in the internal cohesion of the partitions (sub-communities).

The simplest method to search local maxima of the community influence measure is to start from a random division of the network and move one node, at a time, from one side to the other. If the numerical value of P increases, continue with the next node, or if the value does not increase, return the node back and continue with the next node. This procedure is continued until moving any of the nodes does not increase the value of the community influence measure. The first random division will crucially impact the set of local maxima that can be found. More local maxima are searched by starting from a new random division of the network. This is repeated until no more local maxima are found in a reasonable number of trials. Finally, an understanding of the landscape of local maxima can be achieved.

The method assumes that the network is divided into two communities. However the model provides many solutions for the local maxima of Eq. (2). Communities with high rankings according to the value of P in Eq. (2) are candidates for the split of the original community in real-life. Note that this may not be the most probable solution of the community formation process. Later, in Sections 'Zachary's karate club social network' and 'Les Misérables network' we present results for two community measures: the strength P in Eq. (2) of the split into two communities and a statistical measure describing the probability of community formation. In addition, a third measure for robustness of composition of sub-communities is proposed.

In practice, real-world social networks often split into two partitions, although later more sub-communities can build up. However, this kind of community formation is a special case of the basic model. If the original network is first divided into A and B, and later B is divided into B_1 and B_2, usually divisions $(A \cup B_1) \cup B_2$ and $(A \cup B_2) \cup B_1$ are also local maxima of Eq. (2). To be precise, interactions between nodes in one sub-community include also interactions mediated via paths through other communities. In the model, these paths are included whenever both the source node and the target node are inside one community.

2.2 Influence spreading model

In this section, we present the methodology of modelling influence spreading. Influence spreading in a network of people is modelled with nodes describing individuals and links describing connections between neighbouring nodes. Typically, multiple paths between a pair of nodes exist with possible common parts. The mathematical model considers all paths between all node pairs of a network. From this information we construct the influence spreading matrix, or the probability matrix, that describes influence spreading from source nodes to all other nodes in the network. Because of the complex structure of networks, the spreading probability from node A to node B is not equal to the probability from node B to node A. The influence spreading matric is not symmetric, expect in the rare case of a symmetric network topology. One consequence of this is that peripheral nodes that are locally densely connected can have a considerable effect on other parts of the network. Influence spreading accumulates momentum locally at early phases of the process, and later starts a more intensive spreading. [20]

Here, we give an idea how the model can be programmed with a computing language. A more detailed pseudo algorithm has been presented in [20]. The algorithm goes through all paths form a source node to a target node. For every path the probability of spreading is computed by multiplying all the link and node weights along the path. We

denote this factor by W_L where L is the path length (number of links along the path). The interpretation of the link weight is the probability of spreading influence via the link and the interpretation of the node weight is the probability of spreading influence over the node in a time unit. The maximum length of processing paths in computations can be limited by a parameter. Good values of the parameter can be estimated by monitoring accuracy of calculations and computing times as a function of L in W_L.

In spreading processes we have to consider also the temporal distribution of spreading as a function of the path length. We denote the probability of temporal spreading at least via L links as a function of time T by $S_L(T)$. Mathematically, the probability is the survival function $S_L(T) = 1 - F_L(T)$ where $F_L(T)$ is the distribution function of the temporal spreading probability. However, in this section, we consider only equilibrium states for time T approaching infinity. Finally, in a finite network and assuming that node and link weights are not zero, the spreading has reached all nodes. We have for every L

$$\lim_{T\to\infty} S_L(T) = 1.$$

We write an iterative formula for calculating spreading probabilities between nodes in the network. We consider paths between two nodes in the network: a source node and a target node. Paths from source node s to target node t are combined iteratively in the descending order of common path lengths at the beginning of their paths.

$$P_{i,L_{i,2}} = P_{i,L_{\mathrm{i},1}} + P_{i-1,L_{\mathrm{i}-1,2}} - \frac{P_{i,L_{\mathrm{i},1}} P_{i-1,L_{\mathrm{i}-1,2}}}{W_{L_{\mathrm{i},3}}}, i = 1, \dots, N_L - 1. \quad (3)$$

The path length $L_{\mathrm{i},2}$ in iteration i is $L_{i,2} = L_{i,min(L_{i,1},L_{i-1,2})}$ and the common path length of $L_{\mathrm{i},1}$ and $L_{\mathrm{i}-1,2}$ is denoted by $L_{\mathrm{i},3}$. The number of different paths from the source node to the target node is denoted by N_{L}. The iteration starts with two paths with $P_{1,L_{1,1}} = W_{L_{1,1}}$ and $P_{0,L_{0,2}} = W_{L_{0,2}}$ having the longest common path length $L_{1,3}$. If there are more than two with the same common path length, these paths can be processed in any order. In later steps of iteration, combined paths are processed in the same way as the original paths of the network. A numerical example in [20] illustrates the algorithm in practice. The probability of influence spreading between the two nodes is the final result of the algorithm after all the paths have been processed. Denoting the source node by s and the target node by t we have

$$C_{s,t} = P_{N_L-1,\mathrm{L}_{N_L-1},2}.$$

316 In the last step of the iteration the length of the last combined path is $L_{N_L-1,2}$.

317 Un-normalised out-centrality and in-centrality measures for nodes s and t can be defined as

$$C_s^{(out)} = \sum_{t=1}^{N} C_{s,t} \tag{4}$$

318

$$C_t^{(in)} = \sum_{s=1}^{N} C_{s,t}. \tag{5}$$

319 Normalised versions of these centrality measures can be obtained by dividing the expressions by N or $N-1$
320 depending on whether the diagonal elements of influence matrix C are set to one or zero. The corresponding
321 betweenness centrality measure for node n can be defined as

$$b_n = \frac{C - B_n}{C}, \tag{6}$$

322 where

323 $$C = \sum_{s,t=1}^{N} C_{s,t}$$

324 and B_n is calculated similarly to C with node n removed from the network

325 $$B_n = \sum_{\substack{s,t=1 \\ n \notin V}}^{N} C_{s,t}.$$

326 **2.3 Detected communities and their structure**

327 We use two small networks to compare results of the network connectivity and the influence spreading model. We
328 compare the results of the community detection method presented in Section 'Community detection method' Fig. 2
329 shows the general arrangement of our investigations. Two network topologies demonstrate the methods: the Zachary's
330 karate club social network and the Les Misérables network. These networks have been used as example networks in
331 several studies in the literature [22, 23].

332 The two network models have different application areas, definitions and parameterisations. However, for the two
333 example network topologies, the most important communities and their sub-communities are close to each other. More
334 differences appear in weaker communities and in structure. The fact that the same community detection algorithm

provides reasonable results for different network models suggests that the method is generally useful for community detection in various applications. On the other hand, detailed features of the community detection method are important because sub-structures are also discovered. The community detection method is not limited to particular network characteristics: directed, weighted, time dependent, and layered networks can be analysed.

The network connectivity model describes connectivity between pairs of nodes in the network. This is calculated by considering all possible paths between a source node and a target node. We use the same parameter value for describing an operational link between all two neighbouring nodes and utilise the second formula of $r(V)$ in Sect. 3. In social networks low weighting factors are used for describing probabilities of social influence [21]. Although we apply the network model designed for physical communication network modelling, we use low values for link weights as in the modelling of influence spreading.

The main focus of this paper is in the methodology and this is why simple social networks, the Zachary's karate club and the Les Misérables networks are used to demonstrate the method. In addition, we show very detailed results provided by the model in order to demonstrate the granularity and different aspects of the model. However, these results are not analysed in detail because such low-level empirical information is not available. Usually the model predicts the strongest communities accurately but weaker structures are more sensitive to network models and parameter values.

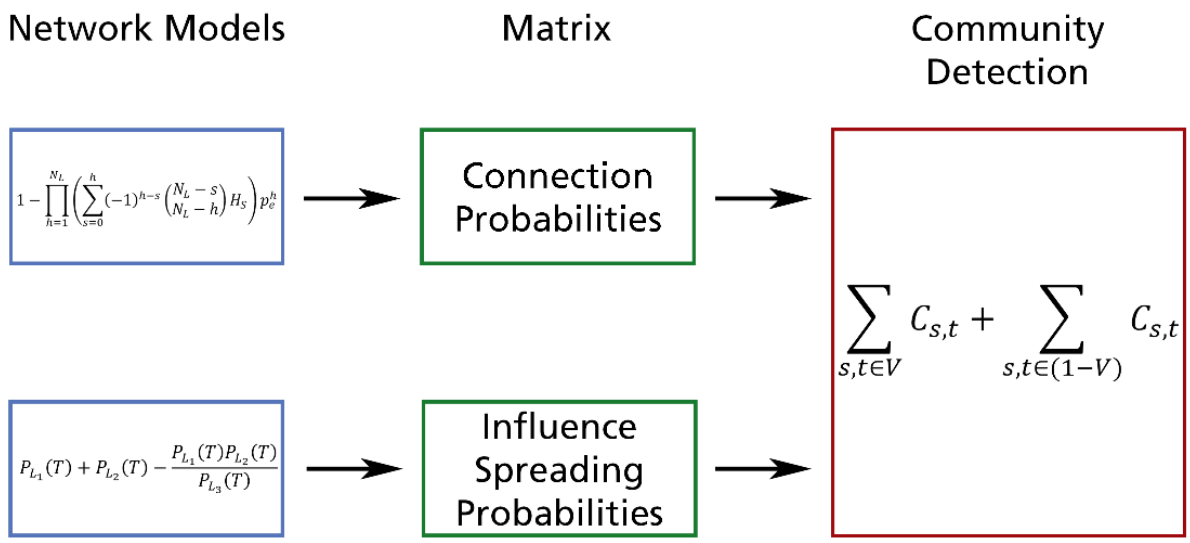


Fig. 2. Two network models and a community detection model are used to demonstrate the methodology.

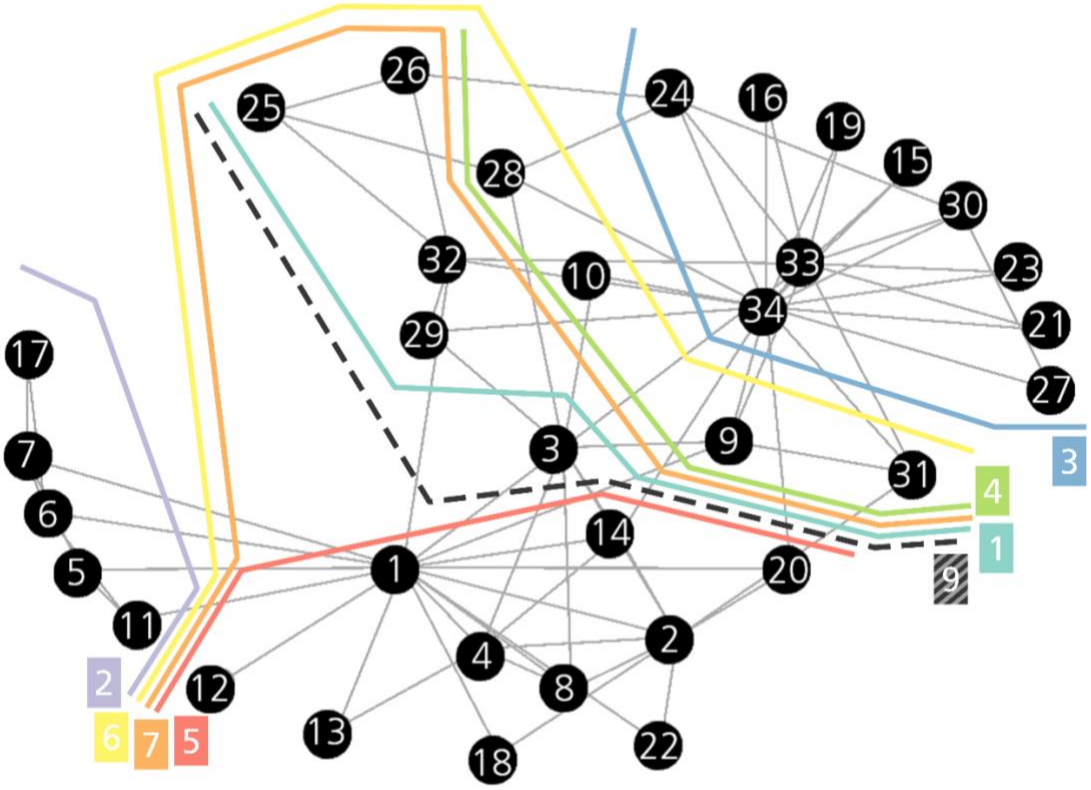
2.3.1 Zachary's karate club social network

W. W. Zachary observed 34 members of a karate club over a period of two years [22]. During the study a disagreement developed between the administrator of the club and the club's instructor. The instructor started a new club, taking 16 members of the original club with him. The karate club social network is pictured in Fig. 3 where line 1 indicates the two partitions after the split of the club with the exception of node 9 who joined the other club. The instructor is node 1 and the administrator is node 34.

2.3.1.1 Community structure

Zachary's karate club is a social network where we assume that low link weights describe the probability of social influence. In addition, only one community, where all nodes of the network are in one community, is detected with high parameter values. This is not an interesting case in our study. Connectivity and influence spreading probabilities describe different phenomena but they can have some common interpretation in social networks. In fact, the two models with the link weight value of 0.05 predict similar rankings for the seven strongest divisions. The compositions of the seven divisions are listed in Table 1. Table 1 gives both partitions of the divisions but sometimes we show only the partition with less nodes in order to simplify notations.

Figure 4 shows the hierarchical structure of the fourteen communities in Fig. 3 and in Table 1. Note that each split has two partitions. Ten different four level structures are detected in the network under communities id2_29, id3_24 and id5_24. These can be verified as, for example, community id1_18: {9-10, 15-16, 19, 21, 23-34} and community id6_19: {5-7, 9-11, 15-17, 19, 21, 23-24, 27-28, 30-31, 33-34} are sub-communities of community id2_29: {1-4, 8-10, 12-16, 18-34}. Two more levels exist below community id1_18 and community id6_19. Colours in Table 3 indicate the two partitions for each seven divisions.



375

376

377 Fig. 3. Zachary's karate club network with divisions indicating detected communities. Divisions 1-7 correspond to
378 lines in Tables 1-3. The additional division 9 is detected in a sensitivity test where link weights have a small random
379 component (see Fig. 8).

380

381 Table 1. The connectivity and influence spreading models with weights 0.05 predict the same rankings for the seven
382 divisions in Fig. 3. Compositions of the split into two partitions and the number of nodes in each partition are
383 shown.

Divisions in Figs. 3 and 4 from connectivity and influence spreading models $(p, w_L = 0.05)$				
id	*#*	*Nodes*	*#*	*Nodes*
id1	16	1-8 11-14 17-18 20 22	18	9-10 15-16 19 21 23-34
id2	5	5-7 11 17	29	1-4 8-10 12-16 18-34
id3	10	15-16 19 21 23-24 27 30 33-34	24	1-14 17-18 20 22 25-26 28-29 31-32

id4	14	9-10 15-16 19 21 23-24 27-28 30-31 33-34	20	1-8 11-14 17-18 20 22 25-26 29 32
id5	10	1-2 4 8 12-14 18 20 22	24	3 5-7 9-11 15-17 19 21 23-34
id6	15	1-4,8,12-14,18,20,22,25-26,29,32	19	5-7 9-11 15-17 19 21 23-24 27-28 30-31 33-34
id7	15	5-7,11,15-17,19,21,23-24,27,30,33-34	19	1-4 8-10 12-14 18 20 22 25-26 28-29 31-32

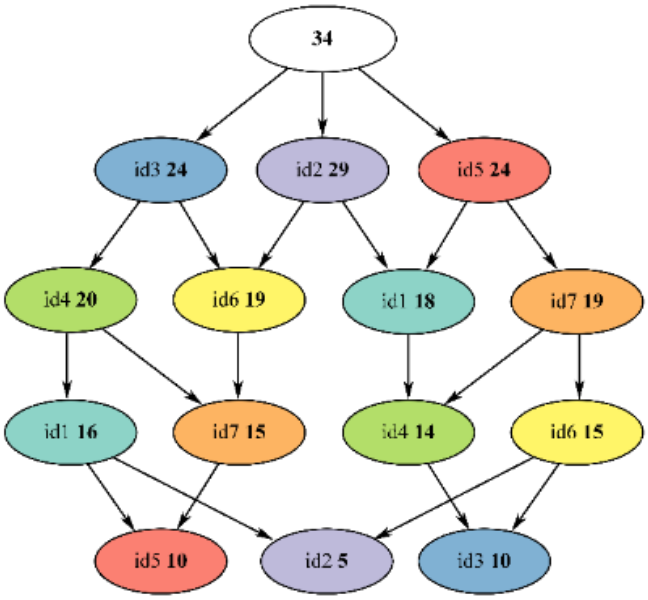


Fig. 4. Hierarchical structure of karate club sub-communities. Seven divisions $id1, \ldots, id7$ detected by the community detection algorithm are indicated by colours (compositions of the 14 communities and sub-communities are listed in Table 1).

Next, we present results of the Zachary's karate club network from the two network models. In Table 2, columns 'A0.05' and 'A0.1' are from the network connectivity model and the other five columns are from the influence spreading model. Nine different solutions for communities are detected. This are lines 1-9 in Tables 2 and 3.

The numerical values of the community influence measure of Eq. (2) from the two network models for the nine detected divisions are shown in the left part of Table 2. The corresponding values of statistical community measures are shown on the middle part of the table. The statistical values are probabilities to split into the two communities. These results are simulated by starting from random initial configurations. The second division in line 2 has the highest

community measure of Eq. (2) for ‘A0.05’ for the network connectivity model and four influence spreading model calculations with different model parameters.

Table 3 shows the nodes included in the communities. For example, the second line indicates that nodes {5, 6, 7, 11 and 17} and {1, 2, 3, 4, 8, 9, 10, 12, ..., 16, 18, …, 34} are members of the two detected communities. The last two columns show that the number of nodes in the communities are 5 and 29. Note, that computer runs ‘A0.1’ and ‘P0.1’ have not found the second community, as can be seen also in Table 3 with the second and the fourth zero in ‘1010111’.

Less communities are found with higher link weights and it is even possible that the strongest community calculated with lower link weights is not found or a new combination of nodes emerges. Comparing lines 1 and 9 in Table 3 reveals that the only difference is node 3 moving to the larger partition. This configuration is the only one detected with the higher weight value of $w_L = 0.1$ in the influence spreading model.

Three columns in Table 2 show calculations from the influence spreading model with three different parameters: ‘PT0.1’ with the time of spreading $T = 0.1$, ‘L0.1’ with the limited path length $L = 1$ ($w_L = 0.1$), and ‘VL0.05’ with the limited number of visits $V = 1$ ($w_L = 0.05$) on a node during the influence spreading. These results agree with the basic calculations of ‘A0.05’ and ‘P0.05’. It is possible that these results are different in other more complex network topologies or network configurations.

Table 2. The values of the community influence measure of Eq. (2) from the two network models for the nine detected divisions are shown in the left part of the table. The corresponding values of the statistical community measures are shown in the middle part of the table. Columns 'A0.05' and 'A0.1' show the results from the network connectivity model with connectivity probabilities $p = 0.05$ and $p = 0.1$ between neighbouring nodes. Columns 'P0.05' and 'P0.1' shows the results from the influence spreading model with influence spreading probabilities $w_L = 0.05$ and $w_L = 0.1$. The next column 'PT0.1' shows the results during the spreading process at time $T = 0.1$ ($w_L = 1.0$) (all the other columns show results for time approaching infinity $T \to \infty$). Column 'L0.1' shows the results with the limited path length $L = 1$ ($w_L = 0.1$). Column 'VL0.05' shows the results with the limited number of visits $V = 1$ on a node during the influence spreading process ($w_L = 0.05, L = 1$). Numerical values in left, middle and right parts of the table are for strength (Eq. (2)), statistical and robustness measures of the nine divisions on lines 1-9. Robustness of communities is discussed in Section 'Robustness of community structure'.

	Strength (Community Measure)								Statistical Community Measure								Robustness of Community						
id	A0.05	A0.1	P0.05	P0.1	PT0.1	L0.1	VL0.05	id	A0.05	A0.1	P0.05	P0.1	PT0.1	L0.1	VL0.05	id	A0.05	A0.1	P0.05	P0.1	PT0.1	L0.1	VL0.05
1	9.18	24.90	9.42		17.20	21.29	8.76	1	12.96 %	10.03 %	13.11 %		14.20 %	14.93 %	14.74 %	1	14.46	34.31	14.79		27.50	33.50	14.25
2	10.35		10.62		19.24	23.89	9.72	2	10.69 %		10.29 %		10.79 %	10.65 %	11.08 %	2	19.14		19.61		35.63	43.89	18.05
3	8.69		8.91		16.23	20.04	8.25	3	4.65 %		5.13 %		5.22 %	5.27 %	4.72 %	3	12.50		12.76		23.61	28.49	12.19
4	8.72	23.58	8.94		16.33	20.12	8.32	4	2.58 %	0.35 %	2.71 %		3.15 %	2.00 %	3.89 %	4	12.61	29.05	12.88		24.00	28.83	12.48
5	8.58		8.79		16.04	19.70	8.16	5	1.68 %		1.64 %		1.87 %	1.78 %	2.20 %	5	12.06		12.27		22.84	27.14	11.84
6	8.03	21.46	8.22		15.03	18.34	7.67	6	0.38 %	0.09 %	0.37 %		0.41 %	0.31 %	0.43 %	6	9.87	20.56	10.00		18.81	21.69	9.89
7	7.94		8.12		14.84	18.10	7.56	7	0.40 %		0.45 %		0.48 %	0.42 %	0.48 %	7	9.49		9.61		18.03	20.72	9.43
8		24.78						8		0.12 %						8		33.86					
9				29.27				9				8.83 %				9				37.67			

Table 3. Nodes in communities corresponding to lines in Table 2. For example, line 2 means that nodes 5, 6, 7, 11, and 17 are members of the first partition. In the next column'1010111' indicates that the division in line 2 is found in computer runs 'A0.05', 'P0.05', 'PT0.1' 'L0.1', and 'VL0.05'. The last two columns show the number of nodes in the two partitions of the network.

id	Nodes	Found in runs	N_1	N_2
1	11111111001111001101010000000000000	1110111	16	18
2	00001110001000001000000000000000000	1010111	5	29
3	00000000000000011001010110010010011	1010111	10	24
4	00000000110000110010101100110110011	1110111	14	20
5	11010001000111000101010000000000000	1010111	10	24

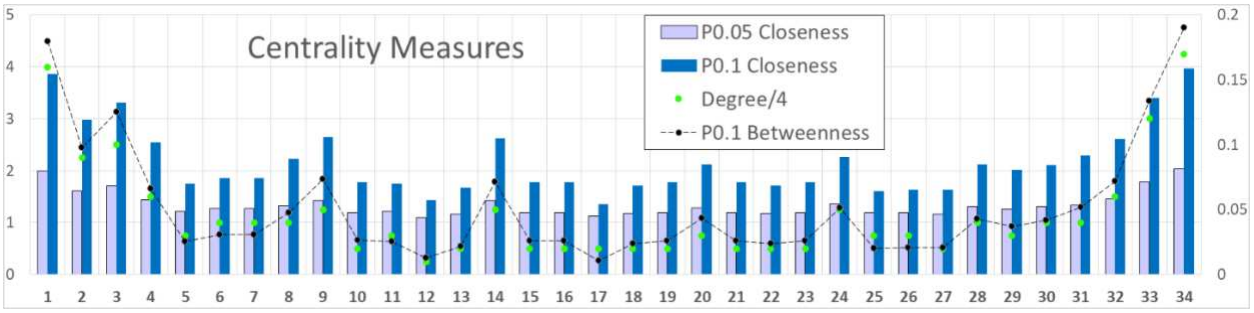
6	11110001000111000101010011001001 00	1110111	15	19
7	00001110001000111010101100100100 11	1010111	15	19
8	11011111001110001100010000000000000	0100000	13	21
9	11011111001111001101010000000000000	0001000	15	19

428

429 **2.3.1.2 Centrality and betweenness results**

430 Figure 5 shows the closeness centrality values of Eq. (4) from the influence spreading model for the 34 nodes of the
431 karate club network. Bars in the figure are for the link weights 0.05 and 0.1. In general, closeness centrality values are
432 higher for higher link weights but some nodes gain some relative advantage, for example, nodes 1, 2, 3, 4, 8, 9, 14,
433 20, 24 and 28-34. These nodes are in central positions (not peripheral) in Fig. 3. Consequently, higher link weight
434 values can strengthen communities in central positions of networks. Degree is a measure of interactions between a
435 node's closest neighbours. For example, nodes 9, 14 and 20 have low degrees but relatively high closeness centrality
436 values. In the network of Fig. 3 these nodes are in central positions. In addition, their betweenness values of Eq. (6)
437 with link weights 0.1 are relatively high. Betweenness curve with link weights 0.05 is not included in the figure
438 because it behaves similarly to the degree curve.

439

440 

441 Fig. 5. Closeness centrality values of Eq. (4) for the 34 karate club members for the influence spreading model with
442 link weights 0.05 and 0.1 are shown with bars. Betweenness centrality values of Eq. (6) for the influence spreading
443 model are shown for link weights 0.1. Node degree values divided by 4 are shown for comparison.

2.3.1.3 Robustness of community structure

Next, we investigate robustness of communities and sub-communities. Robustness for a node belonging to a community is defined as the change of community measure of Eq. (2) when the node is moved from its own partition to the second partition of the division. Because the change in the sum of Eq. (2) is always non-positive, we define the robustness measure as the negative value of change in Eq. (2). Analysing robustness is a method to find nodes who most easily change sides of the main division, division 1 in this case, or other sub-divisions of the network. Strength of divisions and probabilities of division from a random initial state have been presented in Table 2. In the right part of Table 2 robustness values have been shown for the seven computer runs and nine divisions of the karate club network. In this case, strength in the left part of Table 2 has the same ranking of divisions 1-9 as robustness. As strength and robustness are closely related quantities, this kind of results can be expected. However, rankings of the three different community measures can be different due to complex structure of networks. Table 4 shows one example where strength and robustness are not in the same order for the three divisions found in 100 runs with random uniform link weight distribution.

The case of P0.1 from Table 2 has been selected for more detailed analysis because only one division is found with the link weight value of 0.1 in the influence spreading model. The only difference between divisions 1 and 9 in Table 2 is node 3 (line 9 of Table 3 and line 2 of Table 4). Node 3 in division 1 has moved to the second partition when higher link weights are used. In other words, the second partition has more attraction towards node 3. Sensitivity analysis of 100 computer runs with random uniform link weight distribution shows that division 1 with nodes {1-8, 11-14, 17-18, 20, 22} on line 1 in Table 1 also appears in 52 runs. Two weaker divisions in Table 4 are found in 28 and 22 runs.

Figure 6 shows average robustness values for the link weight value of 0.1 (P0.1) and the results of the sensitivity analysis of 100 computer runs (P0.1 $U(0.1,0.01)$). The robustness values for nodes are un-weighted average values over all divisions where the node is a member. In addition, the results for the network connectivity model with probability values 0.1 and the influence spreading model with link weight values 0.05 are shown. Note, that the results are presented for nodes of the network and not for divisions.

Robustness is primarily a measure for studying a node's commitment to its communities but it can be calculated for an individual node as average over all its sub-communities as in Fig. 6. The results in Fig. 6 can be analysed for every node of the network but some observations can be made easily. Nodes 1, 33 and 34 are the most robust nodes

and also nodes 2, 4, 24 and 30 are robust members of their communities. Difference between the network models and parameter values can be analysed node by node. For example, nodes 33 and 34 are relatively more robust members of their community in the influence spreading model than in the connectivity model. Both network models give similar results for the robustness of node 1. Comparing the results of the influence spreading model with link weights 0.05 (P0.05) and 0.1 (P0.1) show similar behaviour. However, it is not so clear from the sensitivity analysis of random link weights in Fig. 6. This is one indication that the sensitivity analysis is useful before detailed conclusions can be made.

Nodes 3, 10 and 20 are examples of loosely bound nodes of their communities thus having low robustness values. These nodes appear in several communities and they jump more easily from one community to another when changing model parameter values or network models. Node 3 is one example of this behaviour as discussed earlier. Nodes with low robustness values have high betweenness values if they are in gateway like positions. This is true for node 3 as can be seen from Fig. 5 but it is not true for node 10, for example. Betweenness and robustness are two different concepts although they are related and correlated in usual network topologies.

Figure 7 goes even deeper in investigating the robustness of communities and their structure. Robustness for selected nodes in the karate club network is shown in Fig. 7 for the seven divisions of Table 2. We study the same case of P0.05 as before. Interesting conclusions can be made about nodes which have high betweenness and low robustness values such as nodes 3 and 9. Node 9 is the one who joined the second partition including node 1. From Fig. 7 we can see that crossing any borders of divisions 1, 4 and 7 are easy for node 9 to change the side. It would have been even easier for node 3 to change side because crossing the border of division 1 has a low influence on the strength of the division.

Results of a sensitivity analysis of 100 computer runs for the influence spreading model with uniformly distributed random link weights between 0.04 and 0.06 are shown in Fig. 8. Altogether 17 divisions are detected and they are listed in the table. Strength, statistics and robustness measures are shown as bars and the number runs where the divisions are found are shown as dots. Statistical and robustness measures are low for divisions 6 and 7. The interpretation is that the probability of formation is low and also their robustness is weak. These two measures are correlated but Fig. 8 illustrates that some divisions can have a low probability of formation but a high robustness value. Division 9 is a good example of a robust division. This composition is uncovered by the sensitivity analysis and we have added this division in Fig. 3 with a dotted line. This is the same split as division 1 with one exception of

node 3. This finding is in good agreement with our previous discussion about node 3. Note that division 9 is detected with higher link weights 0.1 in Table 2 as the only optimal solution of Eq. (2).

Table 4. Three divisions found in 100 computer runs of P0.1 with random uniform distribution $U(0.1,0.01)$ link weights. The last column shows the number runs where the divisions are found.

Nodes (First partition of the division)	*Strength*	*Statistics*	*Robustness*	*Count*
1-8,11-14,17-18,20,22	28.99	9.00	37.61	52
1-2,4-8,11-14,17-18,20,22	28.81	9.07	37.70	28
9-10,15-16,19,21,23-24,27-28,30-31,33-34	27.27	0.24	31.38	22

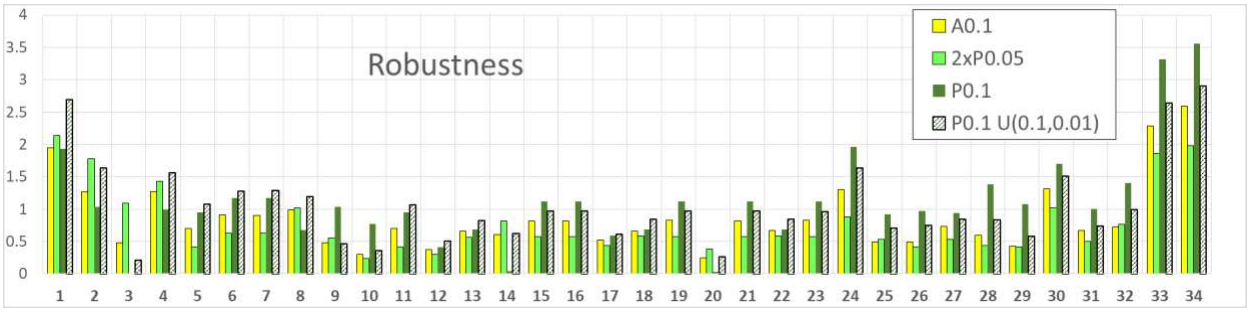


Fig. 6. Robustness values of the 34 karate club members for the influence spreading model with link weights 0.05 (P0.05) and 0.1 (P0.1). Robustness values for the network connection model with probability 0.1 of functioning link (A0.1). For comparison, P0.1 is computed also for uniformly distributed random link weights $U(0.1,0.01)$.

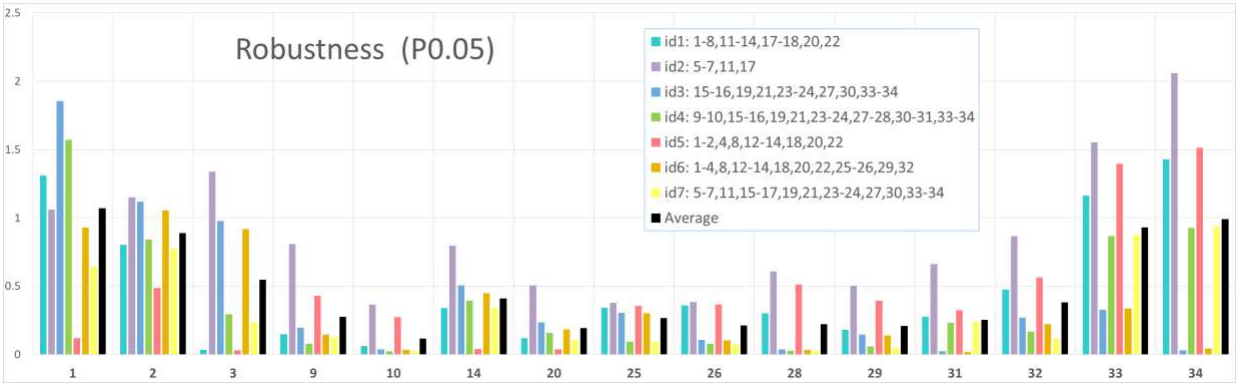
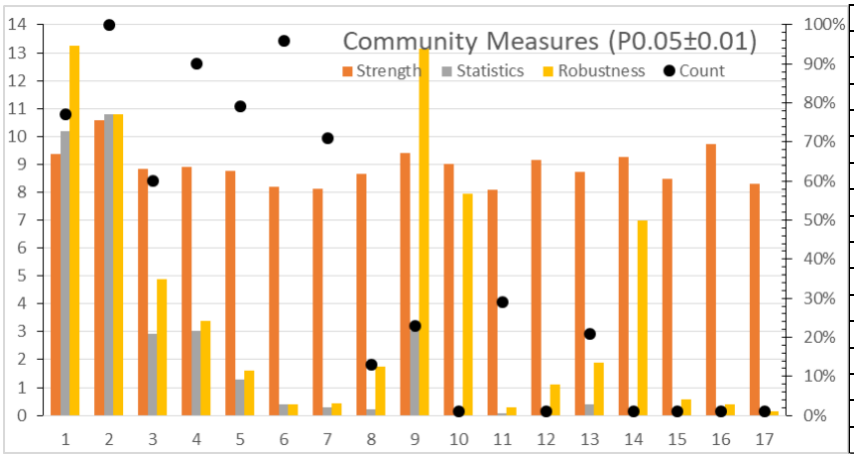


Fig. 7. Robustness values of selected members of the karate club network for P0.05 in Fig. 6. Values are shown for the seven strongest divisions.



1	1-8,11-14,17-18,20,22
2	5-7,11,17
3	15-16,19,21,23-24,27,30,33-34
4	9-10,15-16,19,21,23-24,27-28,30-31,33-34
5	1-2,4,8,12-14,18,20,22
6	1-4,8,12-14,18,20,22,25-26,29,32
7	5-7,11,15-17,19,21,23-24,27,30,33-34
8	1-4,8,12-14,18,20,22
9	1-2,4-8,11-14,17-18,20,22
10	10,15-16,19,21,23-30,32-34
11	5-7,9,11,15-17,19,21,23-24,27,30-31,33-34
12	9,15-16,19,21,23-24,27-28,30-31,33-34
13	9,15-16,19,21,23-24,27,30-31,33-34
14	15-16,19,21,23-30,32-34
15	1-4,8-10,12-14,18,20,22,31
16	1-2,4-8,11-13,17-18,22
17	5-7,10-11,15-17,19,21,23-24,27-28,30,33-34

Fig. 8. Results of a sensitivity analysis from 100 computer runs of the karate club network. The influence spreading model is used with random uniformly distributed $U(0.05, 0.01)$ link weights. Figure illustrates the values of strength, statistical measure, robustness measure and the number of runs where the 17 divisions in the table are detected.

2.3.2 Les Misérables network

Les Misérables is a French novel by Victor Hugo published in 1862. The social network of fictitious characters in the novel have been studied widely in in the community detection literature. Recently the Les Misérables network has been used to study consistency of optimal community structure and an idea that sub-communities correspond to arrangements of a set of underlying building blocks [23]. An information theoretic method was used to discover building blocks from the social network of Les Misérables. Those results can be compared with the findings of this study.

The Les Misérables social network consists of 77 nodes illustrated in Figs. 9 and 10. Figure 9 shows 11 highly ranked divisions from the 201 divisions found in 100 computer runs. The influence spreading model was used with uniformly distributed random link weights between [0.04, 0.06]. Ranking of the 11 divisions is the same for strength of divisions and for probability of formation and very close to the ranking of robustness measure. Figure 11 shows values of the community measures for 50 highly ranked divisions where the 11 divisions of Fig. 9 are the first 11 values.

We propose a new method for discovering building blocks from a network by using borders between divisions provided by a community detection method. Figure 10 shows the corresponding results in [23]. Building blocks discovered in this study are very similar. One difference is that the building block of seven nodes {40, 50, 52, …, 56}

533 is not detected in [23]. In Figure 9 this group of nodes is separated by several boundaries and it can be regarded as a
534 building block. There are other minor differences, for example, nodes {29, 34, 46} are not grouped with nodes {11,
535 14, 15, 16, 33} in the model of this study. Central nodes like 12, 28, 49 and 56 are also members of sub-communities
536 in Fig. 9.

537 Fig. 11 illustrates the community measure values of strength of division, probability of formation (Statistics),
538 robustness of composition and the number computer runs (Count) where the divisions are detected. Compositions of
539 the most highly ranked divisions 1-11 are the corresponding lines in Table 5. Also some divisions 39-46 have good
540 community measure values except that they are only detected in 65% of runs. These divisions are like division 9 in
541 Fig. 8 of the karate club network.

542 Figure 12 illustrates hierarchical structure of the Les Misérables network. We have included only half of the sub-
543 communities detected in the influence spreading model with link weights 0.05. The complete graph of all hierarchical
544 relationships is similar but larger. In Table 5 smaller partitions (sub-communities) are on the left and the graph in Fig.
545 12 shows partitions having node 1 as a member. Figure 12 can be compared with Fig. 4 with the complete hierarchical
546 structure of the karate club network of 34 nodes. Only 6 of the 11 most highly ranked divisions 1, 2, 3, 4, 5 and 8 are
547 detected with link weights 0.05. On the other hand, divisions 12-32, 37 and 49 are found. This is an indication that
548 more sub-communities can be detected by varying link weights.

549 In Fig. 13 robustness of nodes in the Les Misérables network are presented for link weights 0.03 and uniformly
550 distributed random link weights between [0.06, 0.08]. Results are similar except that nodes 17-24 have relatively
551 higher robustness values for lower link weights 0.03. Figure 14 compares closeness centrality, betweenness centrality
552 and robustness of nodes in the Les Misérables network for link weights 0.01. Nodes 12, 28, 49 and 56 are examples
553 of nodes with relatively high betweenness compared to centrality. Robustness of nodes 17-23, 35-39, 58-67 and 77 is
554 relatively high when compared with closeness and betweenness centrality values. These nodes are members of
555 peripheral sub-communities in Fig. 9. Fig. 15 shows a detailed view about the average information of robustness in
556 Fig. 14 for seven strong divisions. Fig. 7 is the corresponding figure for the karate club network.
557

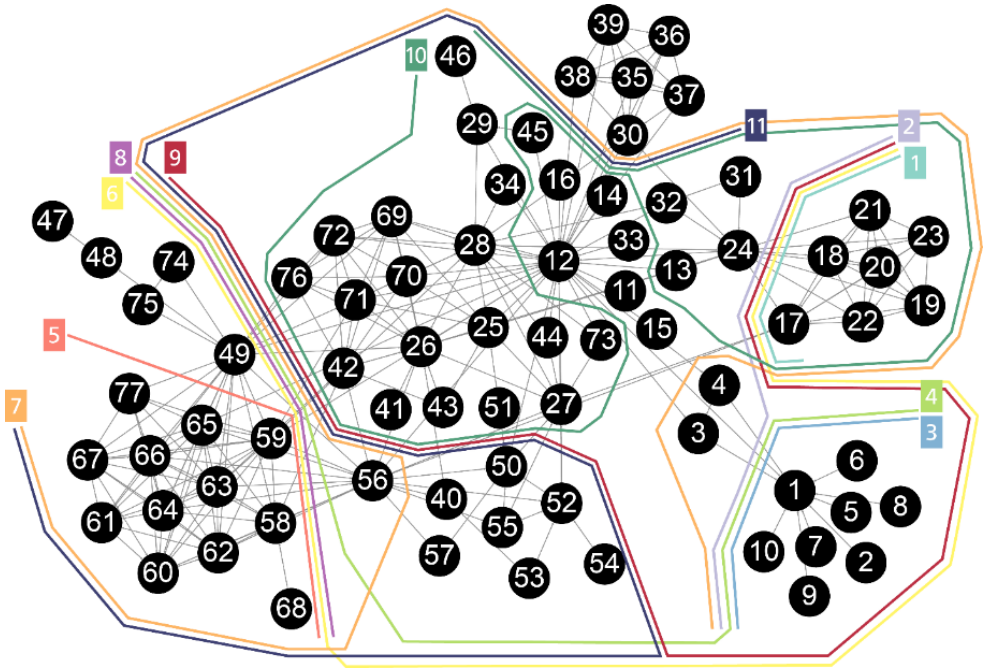


Fig. 9. 11 strong divisions of the Les Misérables network. Borders of the divisions show building blocks of community structure.

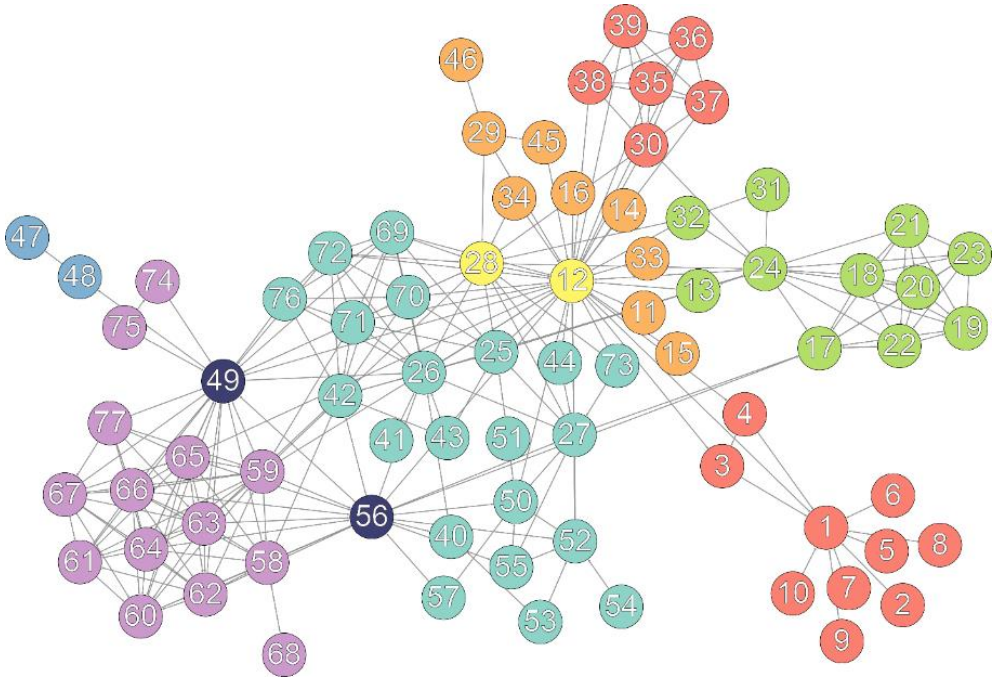


Fig. 10. Building blocks of the Les Misérables network discovered in [23].

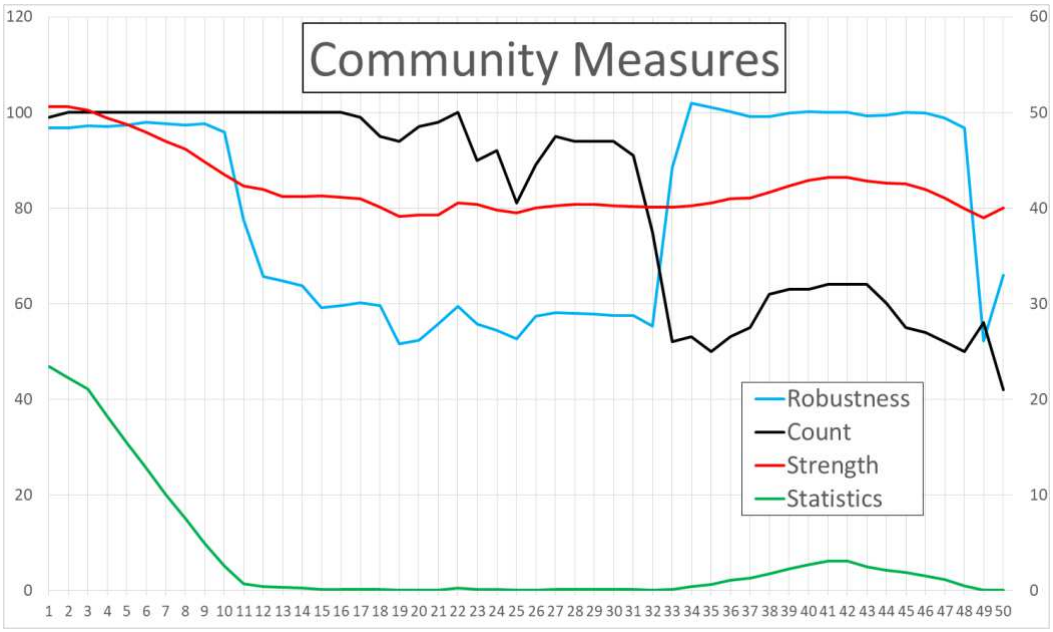


Fig. 11. Values of the community measures for the Les Misérables network. Compositions of the divisions are listed in Table 5. Fifty highly ranked divisions of the 201 division detected in 100 computer runs with uniformly $U(0.05, 0.01)$ distributed random link weights are shown in the figure.

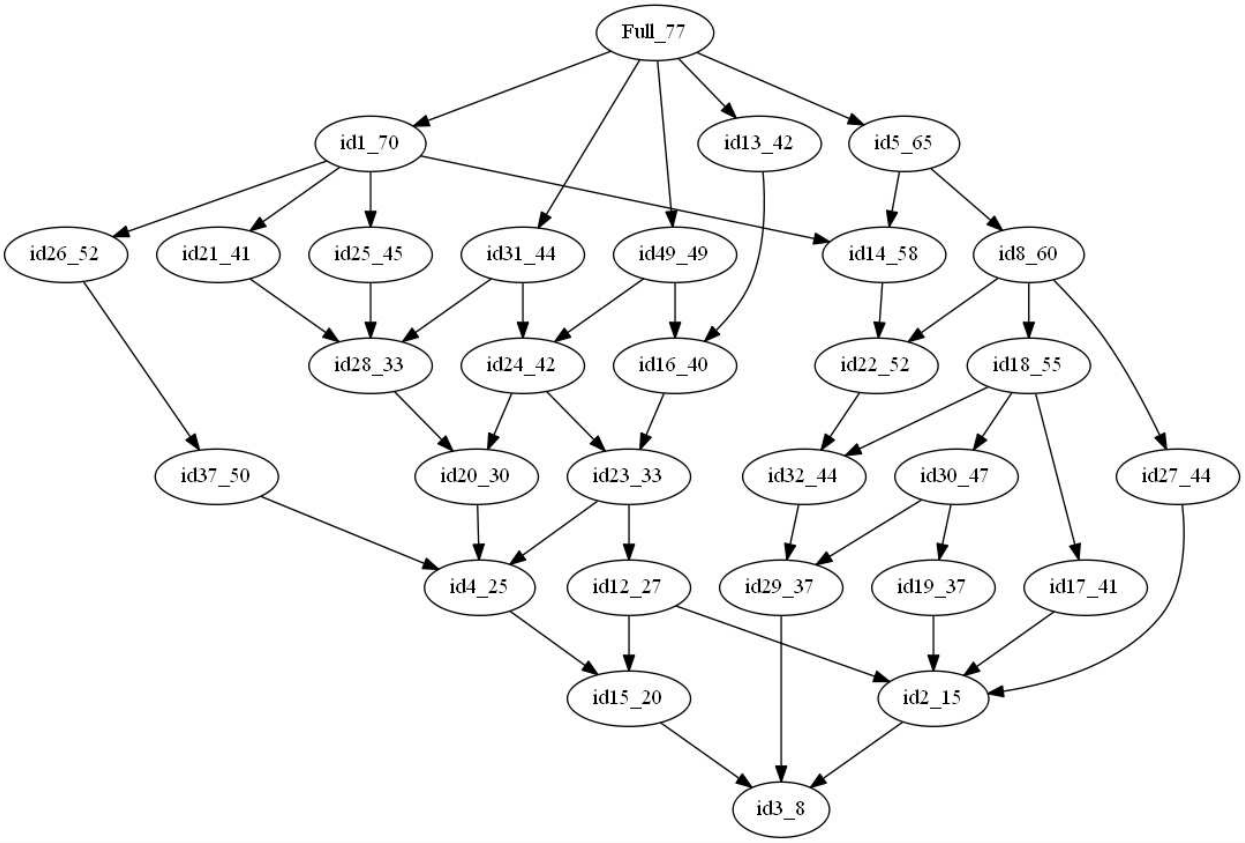


Fig. 12. Hierarchical structure of 29 sub-communities of the Les Misérables network (only half of the hierarchical structure is shown). Divisions indicated by id-numbers are listed in Table 5 with bold ids. Results are for the influence spreading model with link weights 0.05. Sizes of sub-communities are also displayed in the graph.

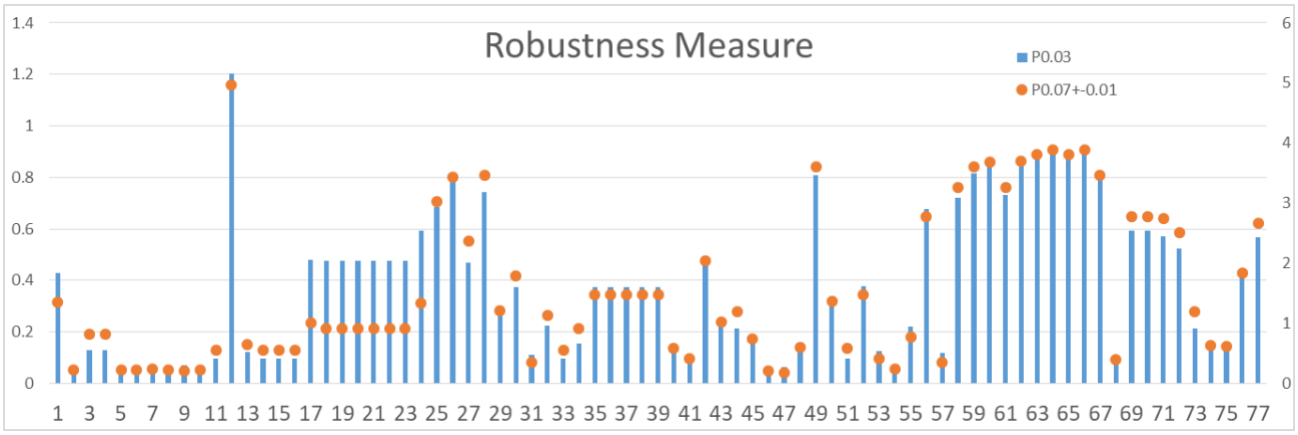
Table 5. Fifty highly ranked divisions out of the 201 detected divisions in 100 computer runs of the Les Misérables network in the influence spreading model with uniformly distributed random link weights $U(0.05, 0.1)$.

id	#	*Nodes*	#	*Nodes*
1	7	17-23	70	1-16,24-77
2	15	1-2,5-10,17-23	62	3-4,11-16,24-77
3	8	1-2,5-10	69	3-4,11-77
4	25	1-2,5-10,47-49,58-68,74-75,77	52	3-4,11-46,50-57,69-73,76
5	12	58-68,77	65	1-57,69-76

6	26	17-23,47-49,56-68,74-75,77	51	1-16,24-46,50-55,69-73,76
7	35	1-10,30,35-39,47-49,56-68,74-75,77	42	11-29,31-34,40-46,50-55,69-73,76
8	17	47-49,58-68,74-75,77	60	1-46,50-57,69-73,76
9	32	17-23,40,47-50,52-68,74-75,77	45	1-16,24-39,41-46,51,69-73,76
10	29	13,17-29,31-32,34,41-44,46,51,69-73,76	48	1-12,14-16,30,33,35-40,45,47-50,52-68,74-75,77
11	31	30,35-40,47-50,52-68,74-75,77	46	1-29,31-34,41-46,51,69-73,76
12	27	1-2,5-10,17-23,58-68,77	50	3-4,11-16,24-57,69-76
13	35	11-16,24-39,41-46,51,69-73,76	42	1-10,17-23,40,47-50,52-68,74-75,77
14	19	17-23,58-68,77	58	1-16,24-57,69-76
15	20	1-2,5-10,58-68,77	57	3-4,11-57,69-76
16	37	3-4,11-16,24-39,41-46,51,69-73,76	40	1-2,5-10,17-23,40,47-50,52-68,74-75,77
17	36	3-4,11-12,14-16,29-30,33,35-39,45-49,56-68,74-75,77	41	1-2,5-10,13,17-28,31-32,34,40-44,50-55,69-73,76
18	22	35-39,47-49,58-68,74-75,77	55	1-34,40-46,50-57,69-73,76
19	37	1-10,13,17-28,31-32,34,41-44,51,69-73,76	40	11-12,14-16,29-30,33,35-40,45-50,52-68,74-75,77
20	30	1-2,5-10,35-39,47-49,58-68,74-75,77	47	3-4,11-34,40-46,50-57,69-73,76
21	36	11-29,31-34,41-46,51,69-73,76	41	1-10,30,35-40,47-50,52-68,74-75,77
22	25	17-23,47-49,56,58-68,74-75,77	52	1-16,24-46,50-55,57,69-73,76
23	33	1-2,5-10,17-23,47-49,56,58-68,74-75,77	44	3-4,11-16,24-46,50-55,57,69-73,76
24	35	3-4,11-16,25-29,32-34,40-46,50-55,69-73,76	42	1-2,5-10,17-24,30-31,35-39,47-49,56-68,74-75,77
25	32	13,17-28,31-32,40-44,50-55,69-73,76	45	1-12,14-16,29-30,33-39,45-49,56-68,74-75,77
26	25	17-26,30-31,35-39,41-43,69-72,76	52	1-16,27-29,32-34,40,44-68,73-75,77
27	33	25-26,28,40-43,47-49,51,53,56-72,74-77	44	1-24,27,29-39,44-46,50,52,54-55,73
28	33	1-10,35-39,47-49,56,58-68,74-75,77	44	11-34,40-46,50-55,57,69-73,76
29	37	1-16,25-29,32-34,41-46,51,69-73,76	40	17-24,30-31,35-40,47-50,52-68,74-75,77
30	30	35-40,47-50,52-68,74-75,77	47	1-34,41-46,51,69-73,76
31	33	11-16,25-29,32-34,40-46,50-55,69-73,76	44	1-10,17-24,30-31,35-39,47-49,56-68,74-75,77
32	33	17-24,30,35-39,47-49,56-68,74-75,77	44	1-16,25-29,31-34,40-46,50-55,69-73,76

33	38	11-12,14-16,30,33,35-40,45,47-50,52-68,74-75,77	39	1-10,13,17-29,31-32,34,41-44,46,51,69-73,76
34	35	1-10,42,47-49,56-72,74-77	42	11-41,43-46,50-55,73
35	37	3-4,11-12,14-16,29-30,33-39,45-49,56-68,74-75,77	40	1-2,5-10,13,17-28,31-32,40-44,50-55,69-73,76
36	33	1-2,5-10,42,47-49,56-72,74-77	44	3-4,11-41,43-46,50-55,73
37	27	13,17-26,30-32,35-39,41-43,69-72,76	50	1-12,14-16,27-29,33-34,40,44-68,73-75,77
38	33	1-2,5-10,17-26,30-31,35-39,41-43,69-72,76	44	3-4,11-16,27-29,32-34,40,44-68,73-75,77
39	36	1-10,17-23,47-49,56-68,74-75,77	41	11-16,24-46,50-55,69-73,76
40	14	1-2,5-10,18-23	63	3-4,11-17,24-77
41	6	18-23	71	1-17,24-77
42	25	42,47-49,56-72,74-77	52	1-41,43-46,50-55,73
43	37	11-34,41-46,51,69-73,76	40	1-10,35-40,47-50,52-68,74-75,77
44	36	3-4,11-24,27,29-39,44-46,50,52,54-55,73	41	1-2,5-10,25-26,28,40-43,47-49,51,53,56-72,74-77
45	38	17-23,30,35-40,47-50,52-68,74-75,77	39	1-16,24-29,31-34,41-46,51,69-73,76
46	36	11-16,24-29,31-34,40-46,50-55,57,69-73,76	41	1-10,17-23,30,35-39,47-49,56,58-68,74-75,77
47	28	13,17-26,28,30-32,35-39,41-43,69-72,76	49	1-12,14-16,27,29,33-34,40,44-68,73-75,77
48	28	11-16,25-29,31-34,41-46,51,69-73,76	49	1-10,17-24,30,35-40,47-50,52-68,74-75,77
49	28	3-4,11-12,14-16,25-29,32-34,41-46,51,69-73,76	49	1-2,5-10,13,17-24,30-31,35-40,47-50,52-68,74-75,77
50	30	25-26,40-43,47-49,56-72,74-77	47	1-24,27-39,44-46,50-55,73

576



577

Fig. 13. Robustness values of nodes in the Les Misérables network from the influence spreading model with link weights 0.03 and uniformly distributed random link weights between [0.06, 0.08].

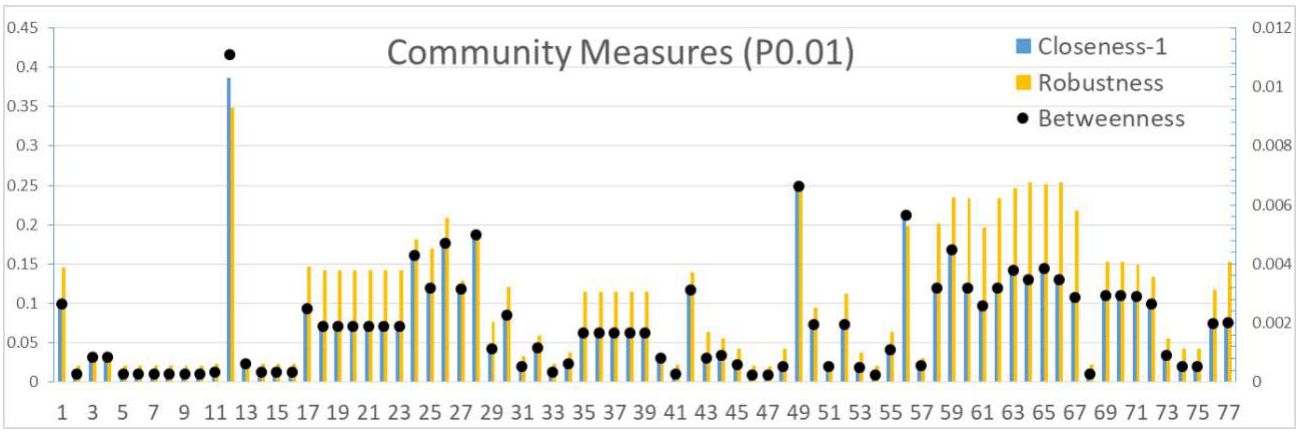


Fig. 14. Closeness and betweenness centrality measures compared with the robustness measure values of nodes in the Les Misérables network. The influence spreading model is used with link weights 0.01. Detailed composition of robustness values are shown in Fig. 15.

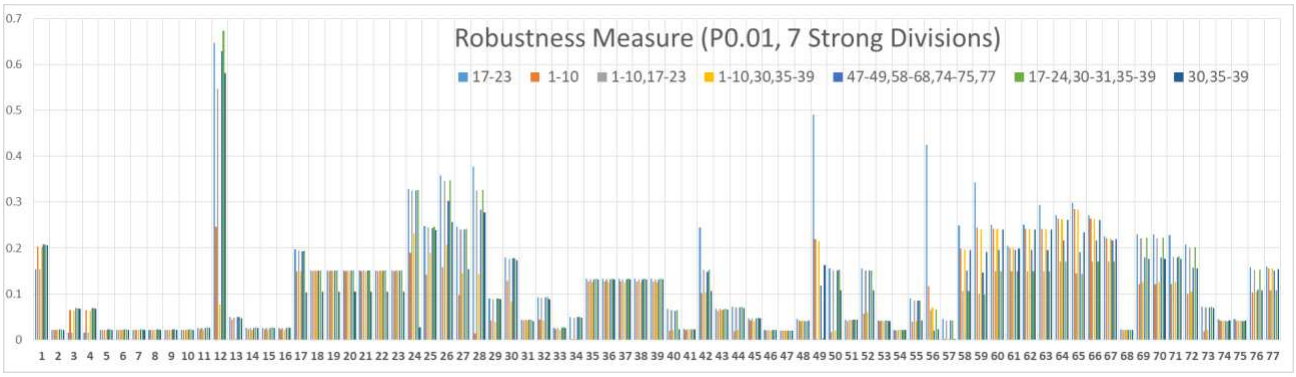


Fig. 15. Robustness values of nodes for seven strong divisions in the Les Misérables network. Corresponding average values of robustness are shown in Fig. 14.

3 Modelling temporal spreading on networks

We assumed in Section 'Influence spreading model' that social networks are in equilibrium state. In the influence spreading model this is accomplished by letting time T approach infinity. This was an appropriate assumption for community detection in situations where social influence has been spreading a long time in the network. In the following sections, we demonstrate how time dependent spreading phenomena can be studied with the model. All centrality and community measures of previous sections can be calculated as functions of time. Community structure is also time dependent. Table 2 shows that similar results for low time values with high link weights and high time values with low link values can be expected. However, in complex network topology there is no simple relation between spreading time and link weights. Also, ranking of sub-communities can change as a function of time. [20]

The focus of Section 'Structure in complex networks' is more structural whereas in this section we study also temporal development. The structural view takes into account topology of the network with link and node weights, and the temporal view studies also time dependent development of spreading processes. Both of the views are built on network topology with node and link weights but in Section 'Structure in complex networks' we have eliminated the time variable by investigating only equilibrium states of networks. In the model, link and node weights along a path of length L are multiplied together and denoted by W_L, and survival functions describing temporal spreading over L links are denoted by $S_L(T)$. Both structural and temporal aspects can be studied by combining these two factors as $W_L S_L(T)$. This holds for a single path between two nodes. The main idea of the spreading model is in the technique of combining multiple paths between node pairs in a network. The method has been presented in Section 'Influence spreading model' and in more detail in [20].

Two different probabilistic distributions are used as examples: the Poisson distribution and the e-mail forwarding distribution. The definition of an e-mail forwarding distribution is provided as a computer algorithm. We demonstrate the model with three empirical social networks: Zachary's karate club [22], a Facebook network [24], and Enron e-mail network [24]. These networks represent small, intermediate and large social networks. The network topology of the Zachary's karate club social network with 34 nodes is pictured in Fig. 3. The Facebook network has 4039 nodes and the Enron e-mail network has 36692 nodes.

Low parameter values for link weights are used for describing influence spreading [20]. Link weights are assumed to be low for one event of an influence attempt because, in a normal social context, the probability of convincing a

person to change his or her opinion is low. For larger networks the influence is expected to be even lower because of less cohesion with more different social groups.

3.1 Poisson and e-mail forwarding survival distribution functions

We study models for time dependency of spreading processes on complex networks [14]. Spreading dynamics and diffusion on networks have studied for example in [15, 25, 26]. In his study, two different temporal spreading distributions demonstrate the modelling with three real-world social networks. The Poisson distribution describes random response time and the e-mail forwarding distribution describes the process of receiving and forwarding messages. Spreading processes are modelled on constant topological network structure. The examples show that the Poisson temporal distribution is more efficient for spreading at low time values for short path lengths and the situation is reversed for high time values and longer path lengths.

In this context, temporal probability distributions describe spreading probabilities starting from a source node via links to other nodes in the network. Survival distribution function $S_L(T)$ provides the probability of temporal spreading via L links at time T. Function $S_L(T)$ is expressed in terms of probability distribution function $F_L(T)$ as $S_L(T) = 1 - F_L(T)$.

The Poisson distribution is a discrete probability distribution that expresses the probability of given number of events occurring in a fixed interval of time if these events occur with a known constant rate and independently of the time since the last event. In this context, the Poisson process describes a process where spreading via successive links occurs randomly as a function of time. For the Poisson distribution $S_L(T)$ is

$$S_L(T) = 1 - \sum_{z=0}^{L-1} e^{-\lambda T} \frac{\lambda T}{z!}, (S_0 = 1).$$

Figure 16 shows Poisson distribution survival functions for path lengths $L = 1, \dots, 20$ and time values $T = 1, \dots, 10$.

The e-mail forwarding distribution describes a typical process of receiving and forwarding e-mails or messages. In the model, the source node sends one e-mail in a time unit and all other nodes in the network check their mail boxes once in a time unit and forward the received e-mail if they had received it before the checking time, or the e-mail stays in the mail box waiting for the next time unit. Nodes send messages and check their mail boxes independently in

642 uniformly distributed random time points. The program code describing the process of receiving an e-mail message
643 once in a time unit and forwarding it once in a time unit is shown below.
644 The two survival distribution functions have different characteristics at short path lengths and at long path lengths.
645 At time $T = 1$ the Poisson probability of reaching nodes at path lengths 1-6 is high compared to the e-mail forwarding
646 probability. Roles are changed with higher path lengths. For example, at time $T = 10$ the spreading probability over
647 path length $L = 20$ for the Poisson distribution is 53 % and for e-mail forwarding 94 %.
648

```
! Survival function S is a function of time t and path length l
S=0
S(1,1)=1.0
do t=1,iT     ! time
  if (t>1) then
    do l=1,iL ! path length
      S(t,l)=S(t-1,l)
    end do
  end if
  do jl=1,iL
    l=iL+1-jl
    do i2=1,l-1
      S(t,l)=S(t,l)+S(t,l-i2)/factorial(i2)-S(t,l)*S(t,l-i2)/factorial(i2)
    end do
  end do
  if (t==1) then
    do l=1,iL-1
      S(t,l)=S(t,l+1)
    end do
  end if
  S(1,iL)=1/factorial(iL)
```

```
end do
```

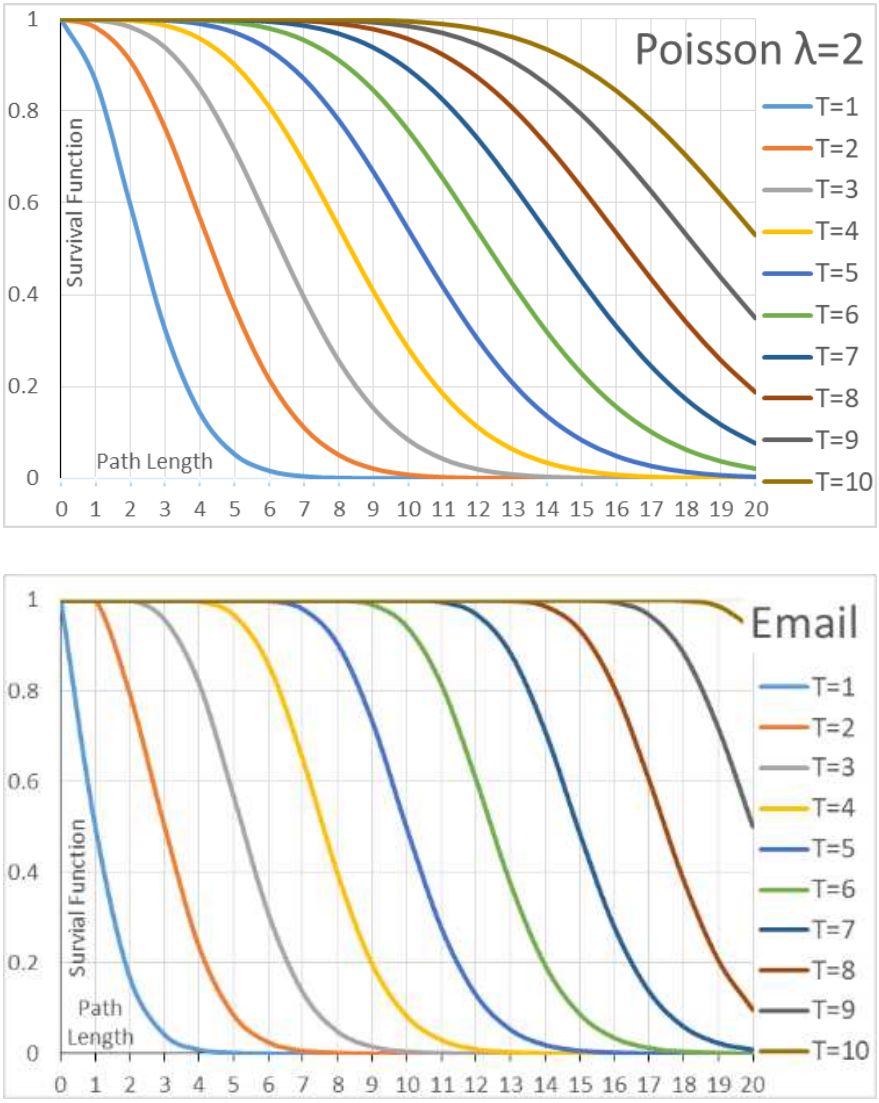


Fig. 16. Survival distribution functions for the Poisson and e-mail forwarding distribution functions as a function of path length for time values $T = 1, \ldots, 10$. The event rate parameter value of $\lambda = 2$ for the Poisson distribution has been used to get comparable results with these path lengths.

Figure 17 shows differences between Poisson and e-mail forwarding survival distribution functions. Curves in Fig. 17 show clearly that Poisson processes dominate e-mail forwarding processes at low time values and short path lengths of $L \leq 10$. Conversely, at high time values and longer path lengths of $L > 10$ e-mail forwarding processes dominate.

Note, that the point on the X-axis, where the roles exchange, depends on the value of Poisson distribution's event rate parameter λ.

These kind of process characteristics can have intricate impacts on how spreading processes perform on complex network topology. At low path lengths the degree of a source node determines how the spreading process starts. The model suggests that Poisson processes are efficient for spreading to neighbouring nodes. On the other hand, spreading processes to distant nodes at high path lengths can get advantage from using e-mail type of delivery methods.

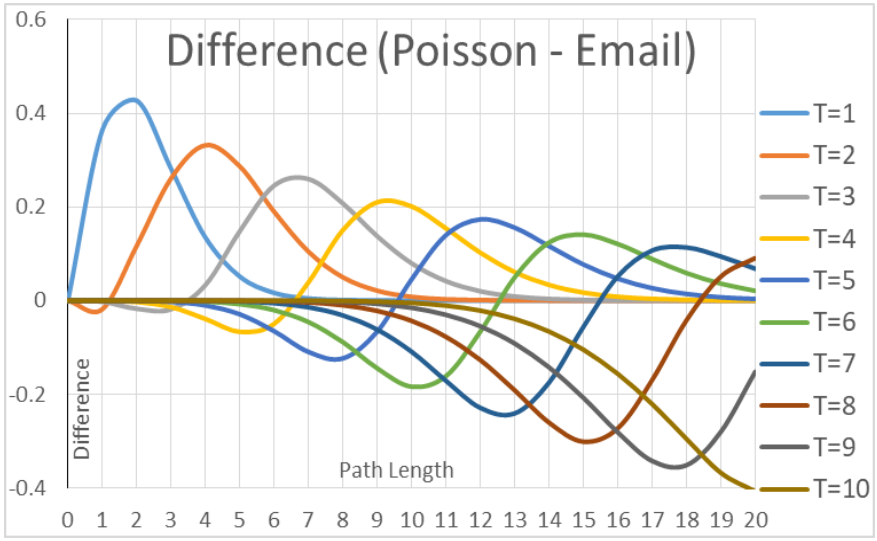


Fig. 17. Difference between Poisson and E-mail forwarding survival distribution functions as a function of path length for time values $T = 1, \dots, 10$.

3.2 Spreading on three empirical social networks

Spreading results of three empirical social networks of the Zachary's karate club network, the Facebook networks and the Enron e-mail network are presented in this section.

Figure 18 shows spreading results for the karate club network with two different link weights $w = 0.05$ and $w = 0.005$. Results are presented for the e-mail forwarding distribution and Poisson distribution with the event rate parameter value of $\lambda = 2$. For both link weights the Poisson spreading process is more effective than the e-mail forwarding process for time values $T < 1.5$. For time $T > 1.5$ the situation is reversed. For higher link weights ($w =$

676 0.05) the spreading process is accelerating for later time values, and for lower values ($w = 0.005$) the process is
677 more linear until the usual saturation effects take over.

678 The expected number of influenced nodes is shown on Y-axes of the figures. This is computed by taking one node,
679 at a time, in the network and assuming that the spreading process is initiated with probability 1.0 at time $T = 0$ from
680 that node. Figs. 18 and 19 show average results over all source nodes in the network. For example, the influence of
681 the Poisson process has spread, on average, to 7.2 nodes with the link weights of $w = 0.05$ at time $T = 1$.

682 Figure 19 shows the expected number of influenced nodes when spreading is initiated from one of nodes 1-34
683 (indicated in the figure) for the Poisson distribution $w = 0.05$. The results agree with the actual situation in the karate
684 club. The instructor of the club is node 1 and the administrator is node 34. In the same figure, the numerical values
685 for the Poisson distribution $w = 0.005$ are show as circles. The expected values of the number of influenced nodes
686 are very low for $w = 0.005$.

687 Rankings for $w = 0.005$ are very close to the corresponding rankings for $w = 0.05$, but they are not exactly the
688 same. For example, node 27 has a higher ranking for $w = 0.05$ than for $w = 0.005$. Node 17 has a relatively higher
689 number of influenced nodes for $w = 0.005$. These are both peripheral nodes but they have different accessibility to
690 central nodes. Node 27 has a link to central node 34 but node 17 has only connections to other peripheral nodes. Node
691 17 has not as much advantage of more active nodes ($w = 0.05$) than node 27. We can see from Fig. 19 that nodes 8,
692 9, 14, 20 have particularly favourable locations in the network. In fact, they have short distances to the most central
693 nodes 1, 33 or 34, and have gateway roles between other nodes in the network. This is a transient phenomenon, because
694 at later time points, influence has propagated more evenly to all nodes in the network.
695

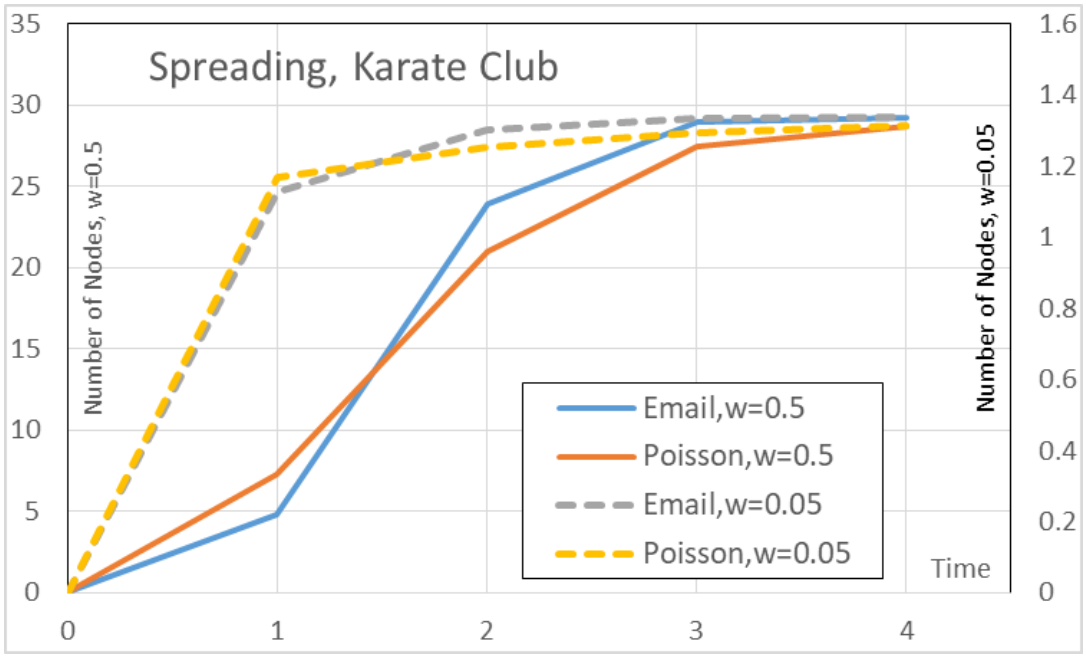


Fig. 18. Spreading in the Zachary's karate club network for the e-mail forwarding process and the Poisson process. Y-axis on the left (right) shows the expected number (normalised) of influenced nodes for $w = 0.5$ ($w = 0.05$).

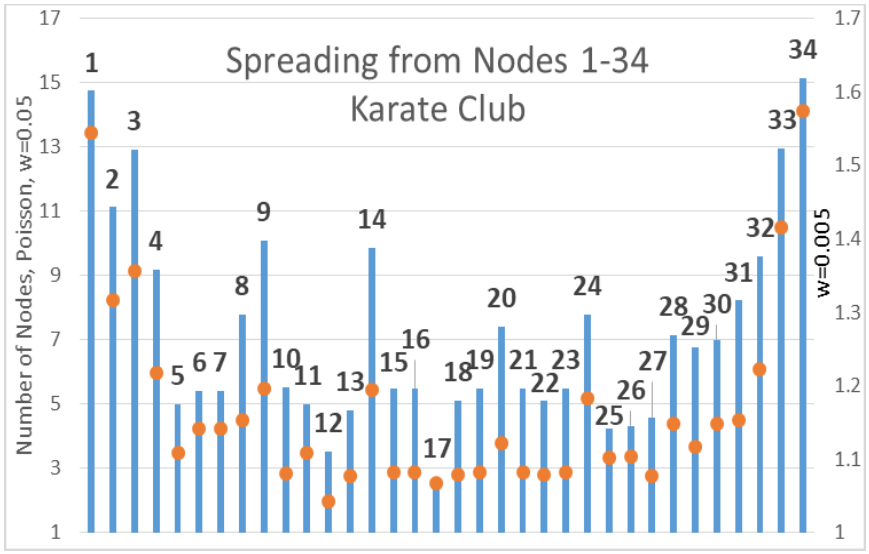


Fig. 19. Spreading from individual nodes 1-34 to other nodes (including the node itself) for the equilibrium state when time approaches infinity. Expected number (normalised) of influenced nodes for $w = 0.05$ on the left Y-axis and for $w = 0.005$ on the right. Numerical values for $w = 0.05$ are shown as bars and for $w = 0.005$ as circles.

The Facebook [24] and Enron networks [24] are examples of larger empirical networks. Figure 20 shows spreading results for these networks with two different link weights $w = 0.005$ and $w = 0.0005$. The results are for total sums from all source nodes because normalised values would be very small. Normalised values can be calculated by dividing the results by the number of nodes in the network. The same phenomenon of crossing curves at time $T = 1.5$ can be seen for both link weights. Again, the Poisson distribution with the event rate parameter value of $\lambda = 2$ is efficient at the beginning of the spreading process and vice versa at later time points. At time $T = 3$ ($w = 0.005$) the e-mail forwarding distribution has reached about 400 nodes more in the Facebook network. With lower link weights ($w = 0.0005$) the difference is less because of weaker spreading power.

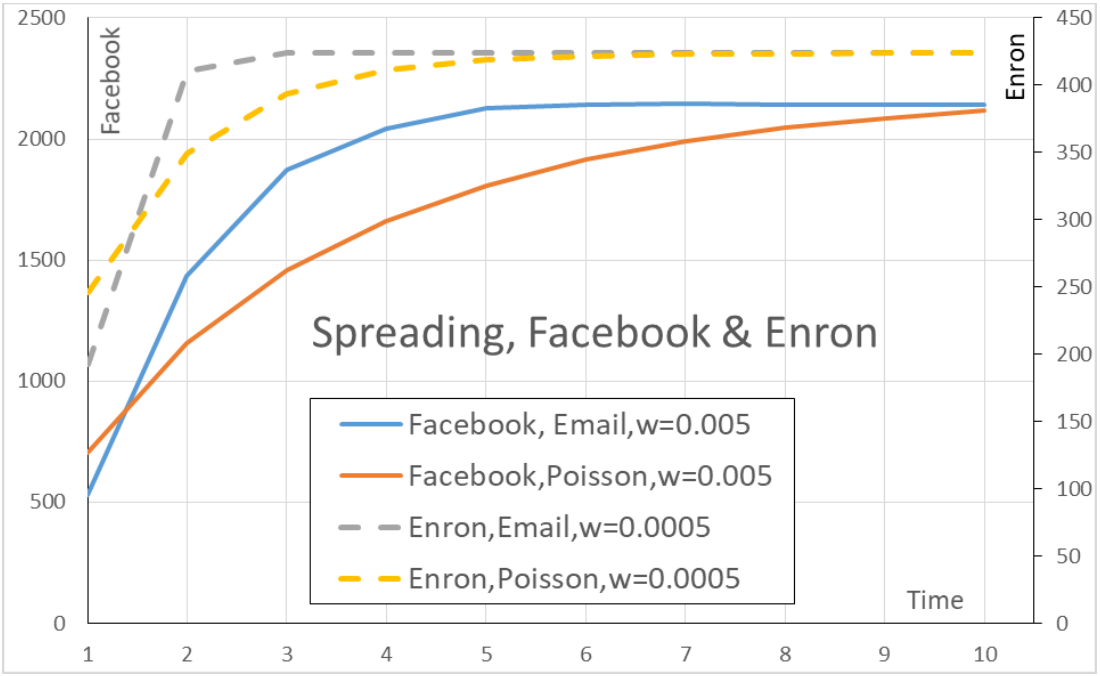


Fig. 20. Spreading in the Facebook and Enron networks for the Poisson process and the e-mail forwarding process. Y-axis on the left (right) shows the expected number (un-normalised) of influenced nodes for the Facebook network (Enron network).

4 Conclusions

We propose a general methodology for analysing complex networks. Models of this paper can be used to study community or module structure and temporal spreading in social, biological and technological networks. The

community detection method is based separating the network model from the community detection model. Different network models can be used to provide input for the community detection algorithm. We demonstrate this by the standard network connectivity model and an influence spreading model. Different temporal distributions can be incorporated into the spreading model simply by using a desired probabilistic survival distribution function.

Communities and sub-communities are identified by local maxima of a quality function that measures the internal strength of two partitions in the network. The quality function used in this study is the sum of probability matrix element values describing interactions between pairs of nodes in both of the two partitions. Cross terms are not included but the model allows influence via paths that go through the other partition. Weak interactions between nodes produce more solutions than strongly connected networks. By varying link weights it is possible to get understanding of the landscape of local maxima and to identify structure in networks. As a new application we propose that the community detection method of this study can be used as a tool for discovering underlying groups of nodes that are usually found together in community structures. Communities and sub-communities correspond to different combinations of these building blocks.

We present different approaches for measuring and ranking communities. This is an important topic because the method usually gives several solutions and highly ranked solutions are candidates for real-life communities. Three different quality measures are presented for the strength of a community, for the probability of formation of a community and for the robustness of a community. These measures are correlated but they represent different views of evaluating communities.

We demonstrate the use of temporal spreading distributions with the Poisson distribution and an e-mail forwarding distribution. The e-mail forwarding distribution is defined with a program code. The Poisson distribution is more efficient for spreading at low time values with short path lengths and the situation is reversed for higher time values and longer path lengths. The exact time value where the crossing occurs depends on the event rate parameter value of the Poisson distribution.

References

1. Barabási A-L: Network Science. Cambridge University Press (2016).
2. Newman MEJ: Networks, An Introduction. Oxford University Press, Oxford (2010).
3. Barrat A, Barthélemy M, Vespignani A: Dynamical Processes on Complex Networks. Cambridge University Press (2008).

4. Coscia M, Giannotti F, Pedreschi D. A classification for community discovery methods in complex networks. Stat. Anal. Data Min. 4(5), 512-546 (2011).
5. Fortunato S, Hric D: Community detection in networks: a user guide. Phys. Rep. 659(11), 1-44 (2016).
6. Lancichinetti A, Fortunato S: Community detection algorithms: a comparative analysis. Phys. Rev. E. 80, 056117 (2009).
7. Yang Z, Algesheimer R, Tessone CJ: A comparative analysis of community detection algorithms on artificial networks. Sci. Rep. 6,30750 (2016). https://doi.org/10.1038/srep30750.
8. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E: Fast unfolding of communities in large networks. Journal of Statistical Mechanics. P10008 (2008)
9. Rosvall J, Bergstrom CT: Maps of random walks on complex networks reveal community structure. PNAS105: 1118 (2008).
10. Karrer B, Newman MEJ: Stochastic blockmodels and community structure in networks. Phys. Rev. E. 83(1), 016107 (2011).
11. Lancichinetti, A, Fortunato, S, Kertész, J: Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics. 11 (2009).
12. Smilkov D, Kocarev L: Influence of the network topology on epidemic spreading. Phys. Rev. E. 85, 016114 (2012).
13. Centola D: How behavior spreads, the science of complex contagions. Princeton University Press (2018).
14. Holme P, Saramäki J: Temporal networks. Phys. Rep. 519, 97-125 (2012).
15. Wang W, Liu Q-H, Liang J, Hu Y, Zhou T: Coevolution spreading in complex networks. Phys. Rep. 820, 1-51 (2019).
16. Gómez S: Centrality in networks: finding the most important nodes. In: Moscato, P, de Vries, NJ (eds.) Business and Consumer Analytics: New ideas. Part III, Chapter 8, pp. 401-434 (2019).
17. Katz L: A new status index derived from sociometric analysis. Psychometrika 18(1), 39-43 (1953).
18. Ball MO, Colbourn CJ, Provan JS: Network Reliability. In: Handbooks in Operations Research and Management Science. Chapter 11. vol 7, pp. 673-762 (1995).
19. 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. COMPLEX NETWORKS 2017. Studies in computational intelligence, vol. 689. Cham: Springer, pp. 202-215 (2018).
20. Kuikka V: Influence spreading model used to analyse social networks and detect Sub-communities. Computational Social Networks 5:12. (2018). https://doi.org/10.1186/s40649-018-0060-z.
21. Kuikka V: A General method for detecting community structures in complex networks. In: Cherifi H. et. al. (eds.) Complex Networks & Their Applications VIII. COMPLEX NETWORKS 2019. Studies in Computational Intelligence, vol. 881. Springer (2019). https://doi.org/10.1007/978-3-030-36687-2_19.
22. Zachary WW: An information flow model for conflict and fission in small groups. J. Anthropol. Res. 33, 452-473 (1977).

778 23. Riolo MA, Newman MEJ: Consistency of community structure in complex networks. Submitted to Phys. Rev. E. (2020),
779 https://arxiv.org/abs/1908.09867.
780 24. Leskovec J, Krevl A: SNAP Datasets, Stanford Large Network Dataset Collection (2014).
781 25. Zhang Z-K, Liu C, Zhan X-X, Lu X, Zhang C-X: Dynamics of information diffusion and its applications on complex networks.
782 Phys. Rep. 651, 1-34 (2017).
783 26. Horváth DX, Kertész J: Spreading dynamics on networks: the role of burstiness, topology and non-stationarity. New. J. Phys.
784 16 073037 (2014).

Figures

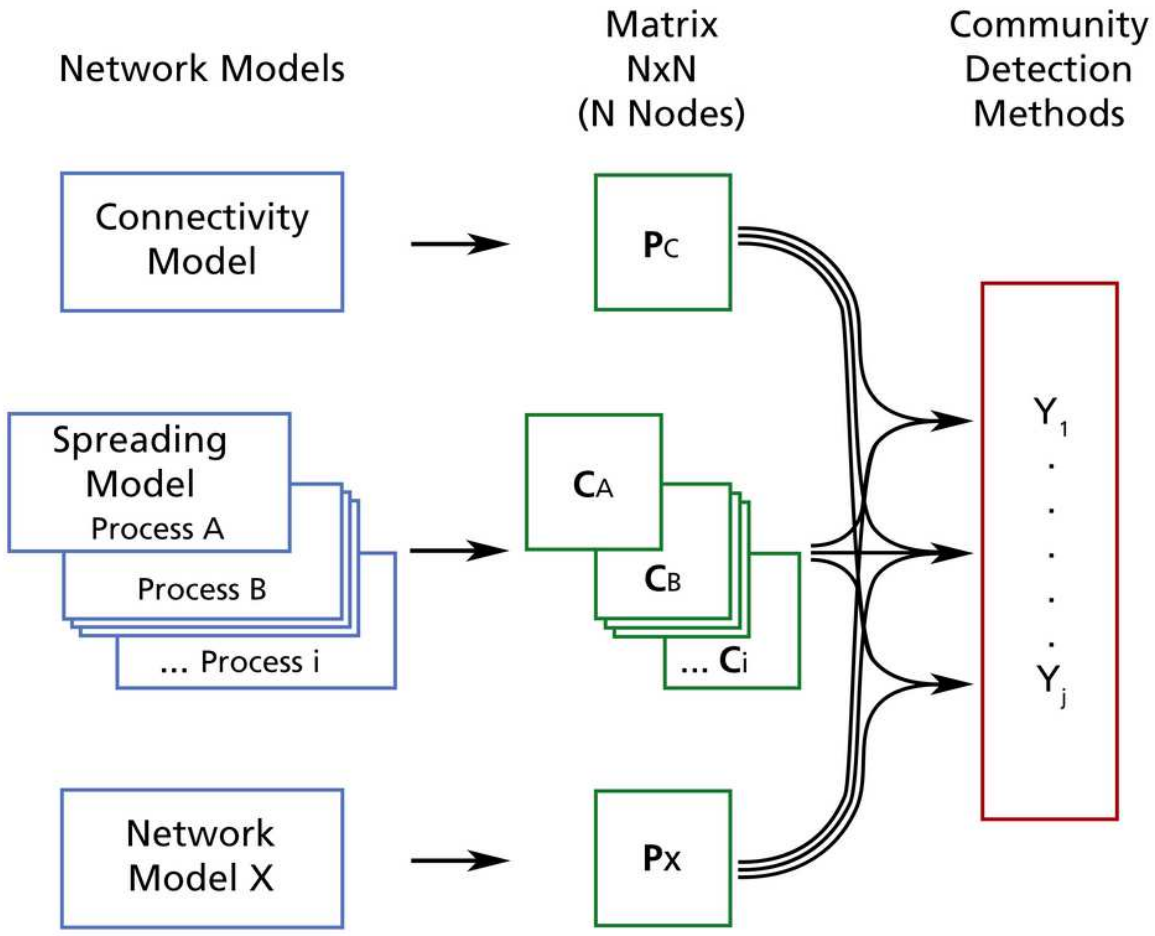


Figure 1

General procedure of using different network models and community detection models. Information from network models are mediated with a probability matrix describing influence or connectivity between all node pairs in the network.

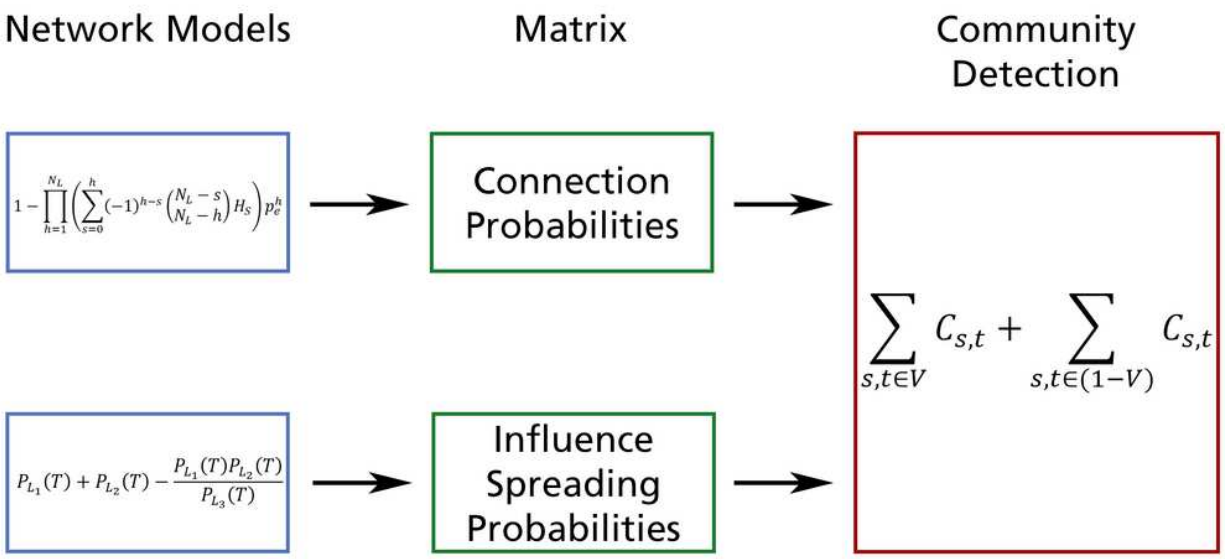


Figure 2

Two network models and a community detection model are used to demonstrate the methodology.

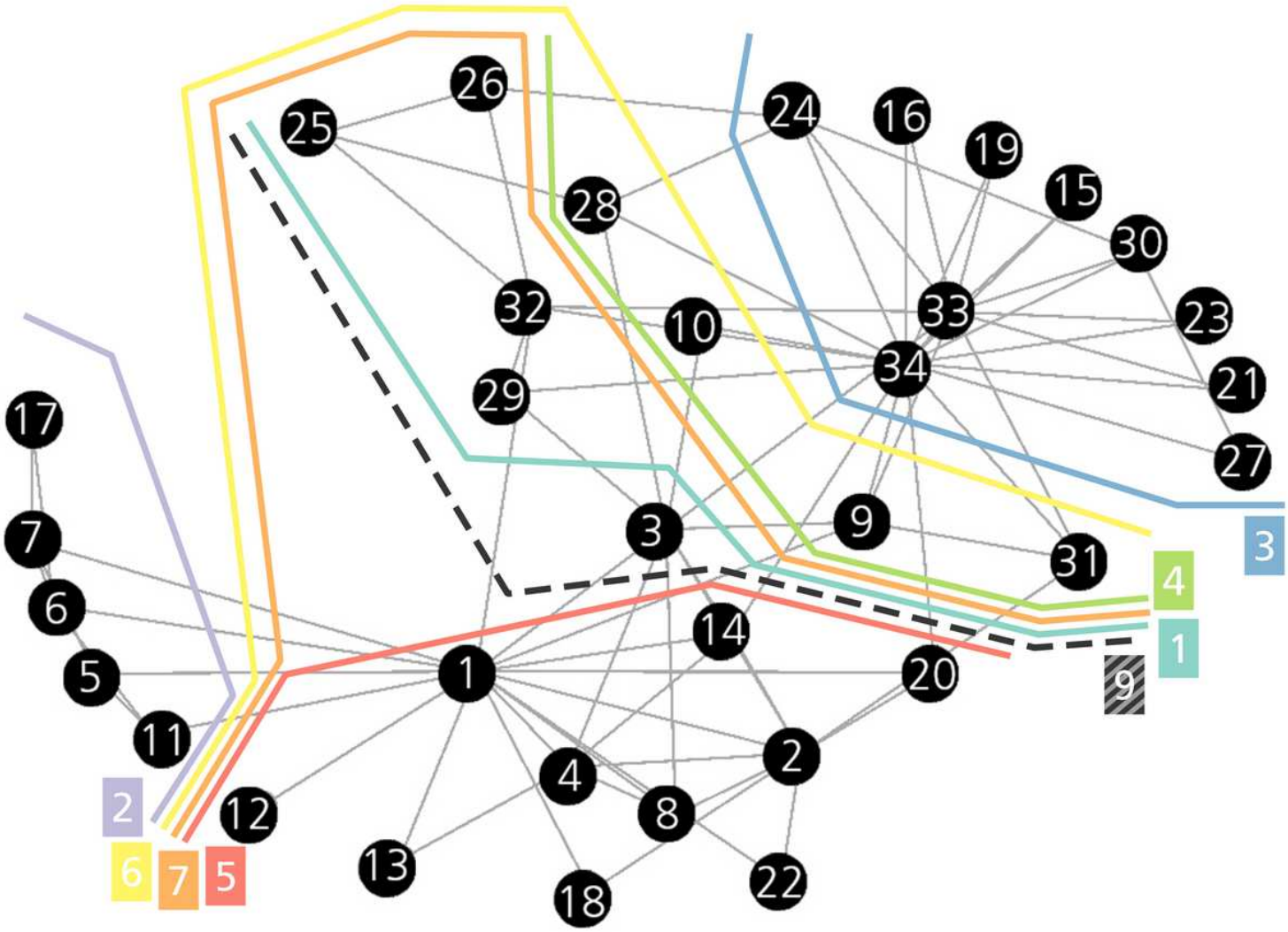


Figure 3

Zachary's karate club network with divisions indicating detected communities. Divisions 1-7 correspond to lines in Tables 1-3. The additional division 9 is detected in a sensitivity test where link weights have a small random component (see Fig. 8).

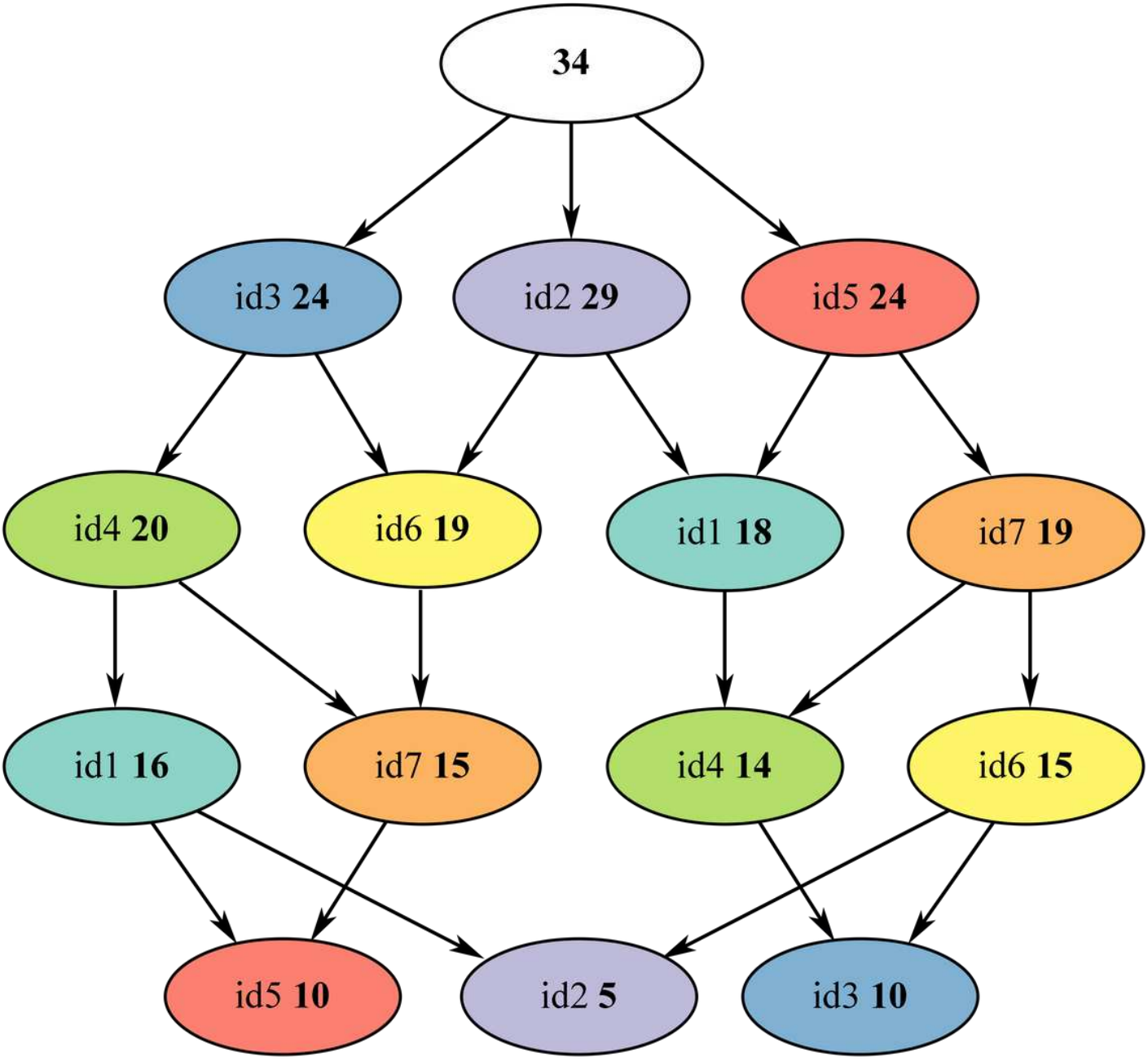


Figure 4

Hierarchical structure of karate club sub-communities. Seven divisions id1,…,id7 detected by the community detection algorithm are indicated by colours (compositions of the 14 communities and sub-communities are listed in Table 1).

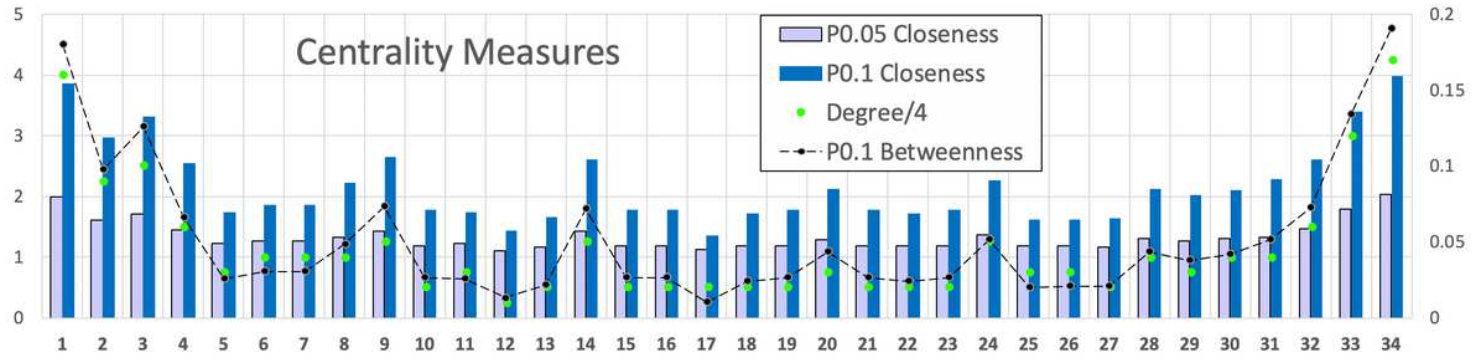


Figure 5

Closeness centrality values of Eq. (4) for the 34 karate club members for the influence spreading model with link weights 0.05 and 0.1 are shown with bars. Betweenness centrality values of Eq. (6) for the influence spreading model are shown for link weights 0.1. Node degree values divided by 4 are shown for comparison.

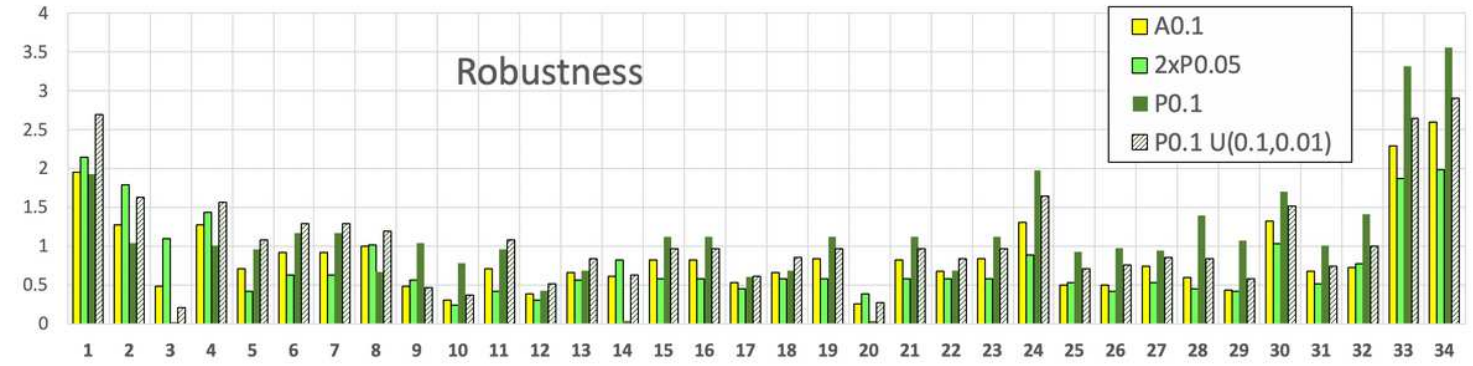


Figure 6

Robustness values of the 34 karate club members for the influence spreading model with link weights 0.05 (P0.05) and 0.1 (P0.1). Robustness values for the network connection model with probability 0.1 of functioning link (A0.1). For comparison, P0.1 is computed also for uniformly distributed random link weights U(0.1,0.01).

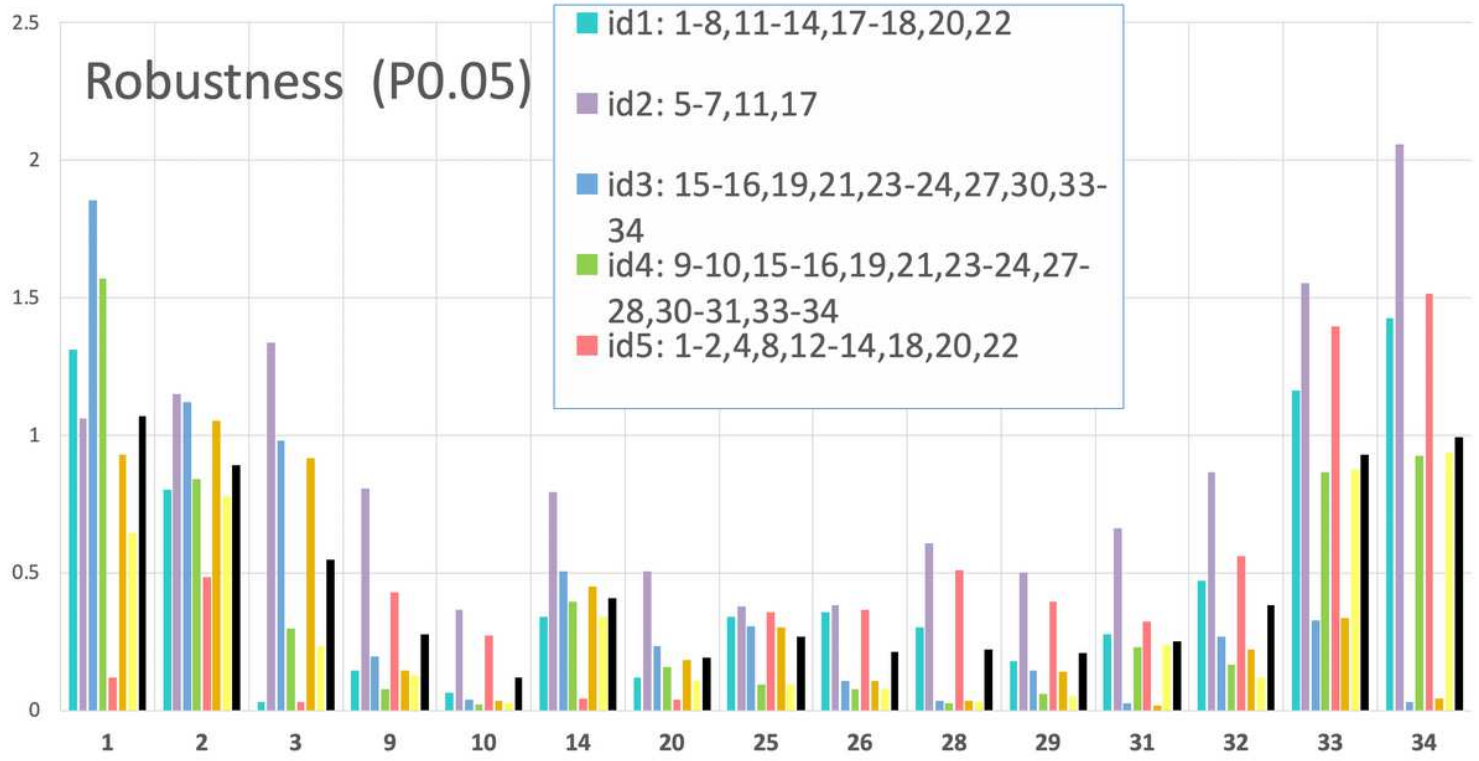


Figure 7

Robustness values of selected members of the karate club network for P0.05 in Fig. 6. Values are shown for the seven strongest divisions.

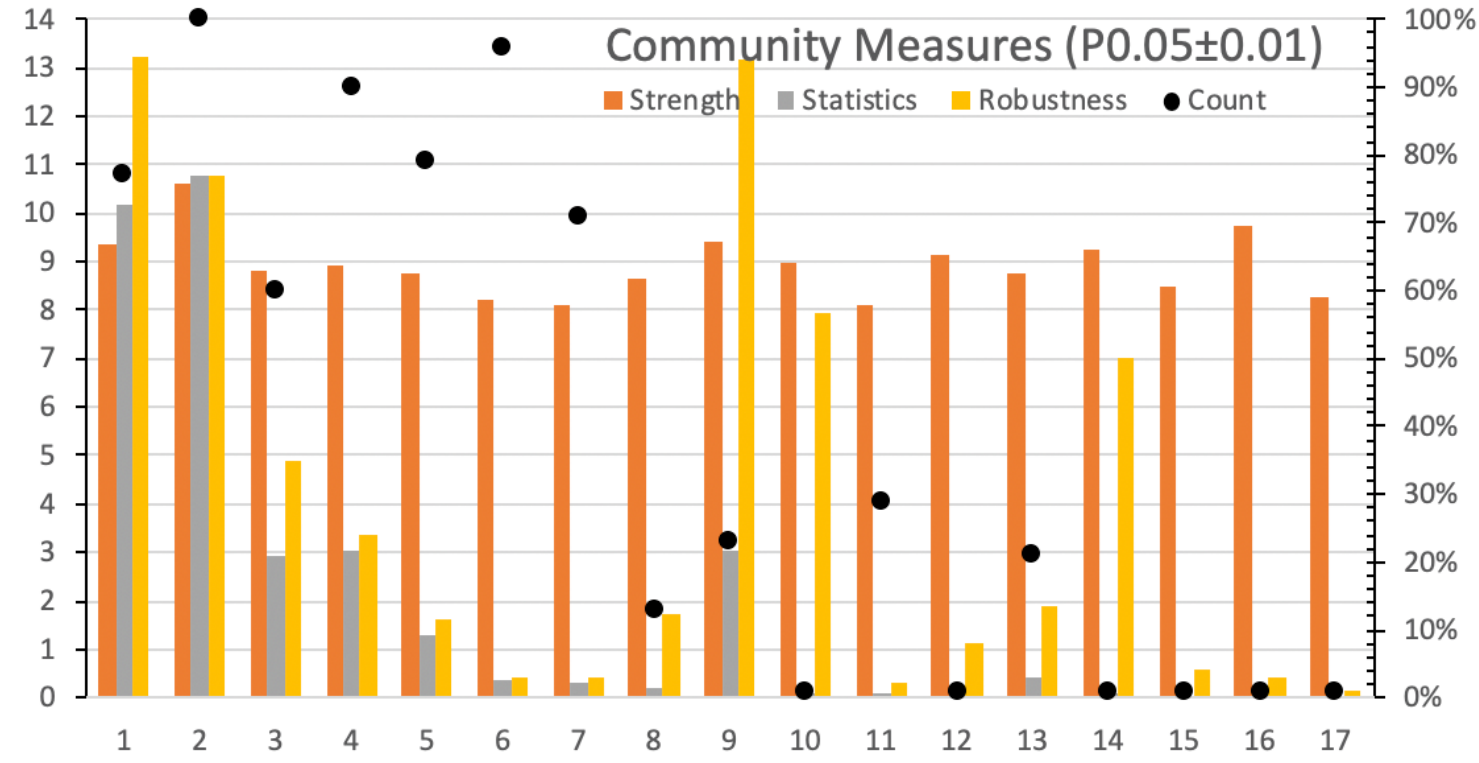


Figure 8

Results of a sensitivity analysis from 100 computer runs of the karate club network. The influence spreading model is used with random uniformly distributed U(0.05,0.01) link weights. Figure illustrates the values of strength, statistical measure, robustness measure and the number of runs where the 17 divisions in the table are detected.

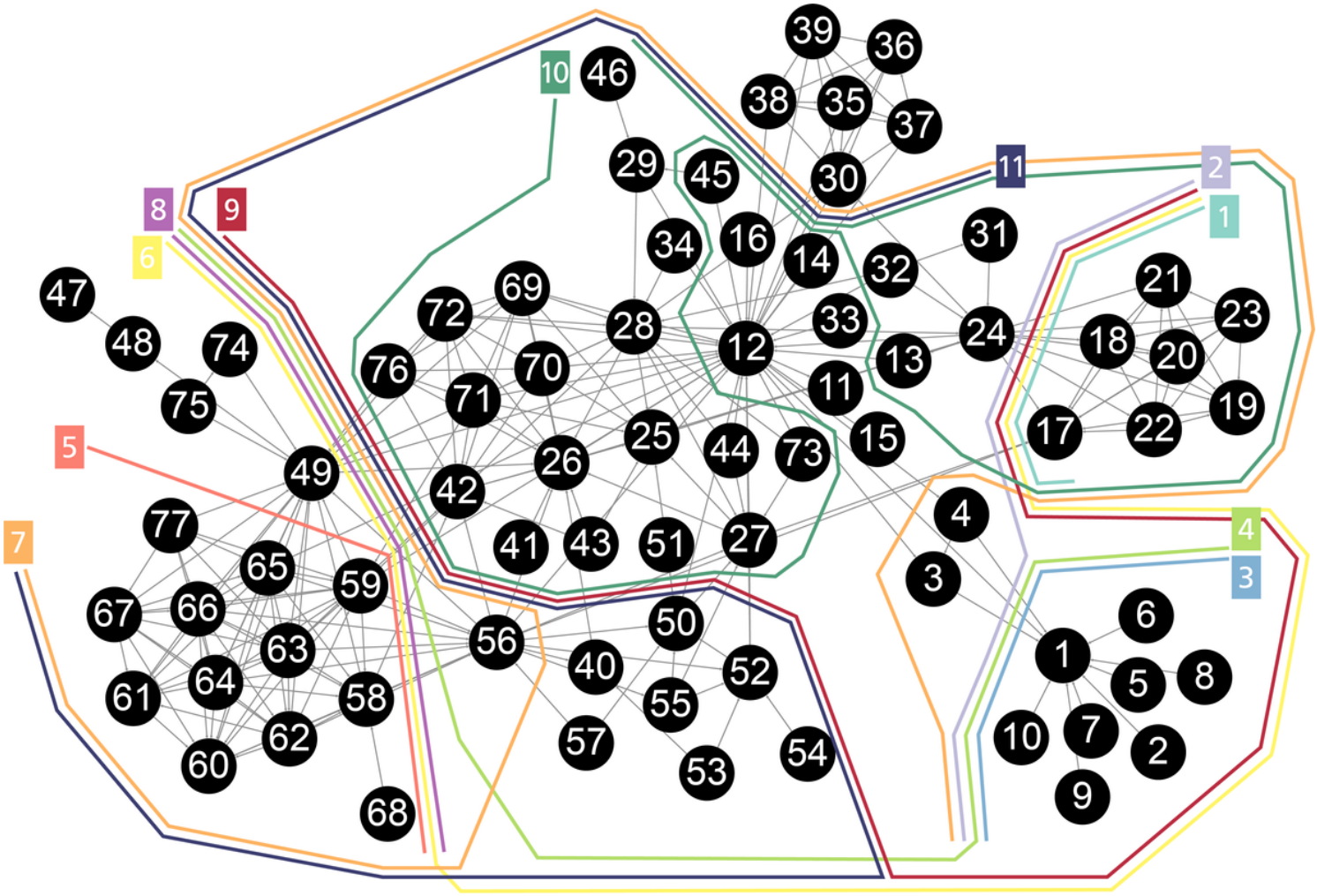


Figure 9

11 strong divisions of the Les Misérables network. Borders of the divisions show building blocks of community structure.

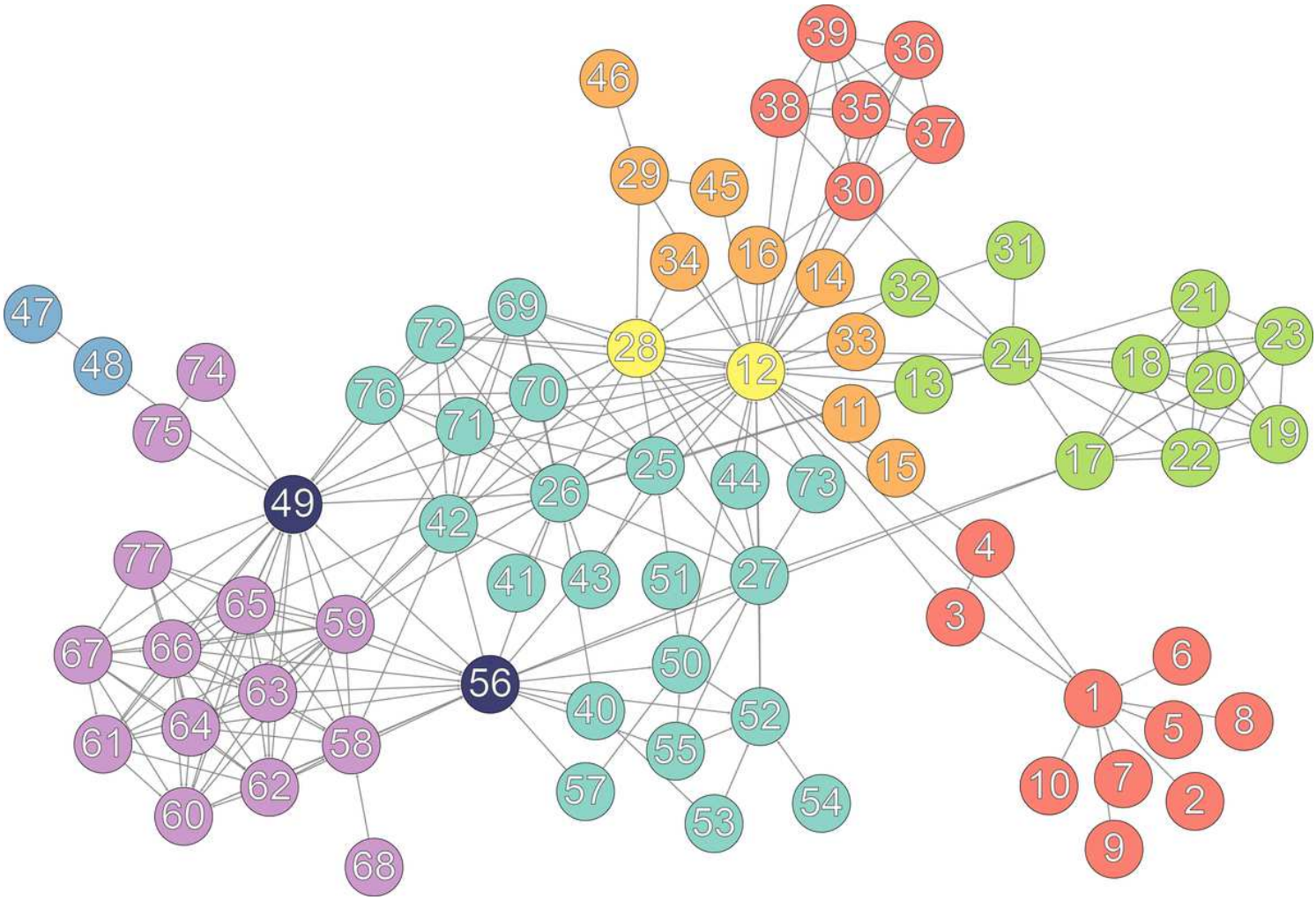


Figure 10

Building blocks of the Les Misérables network discovered in [23].

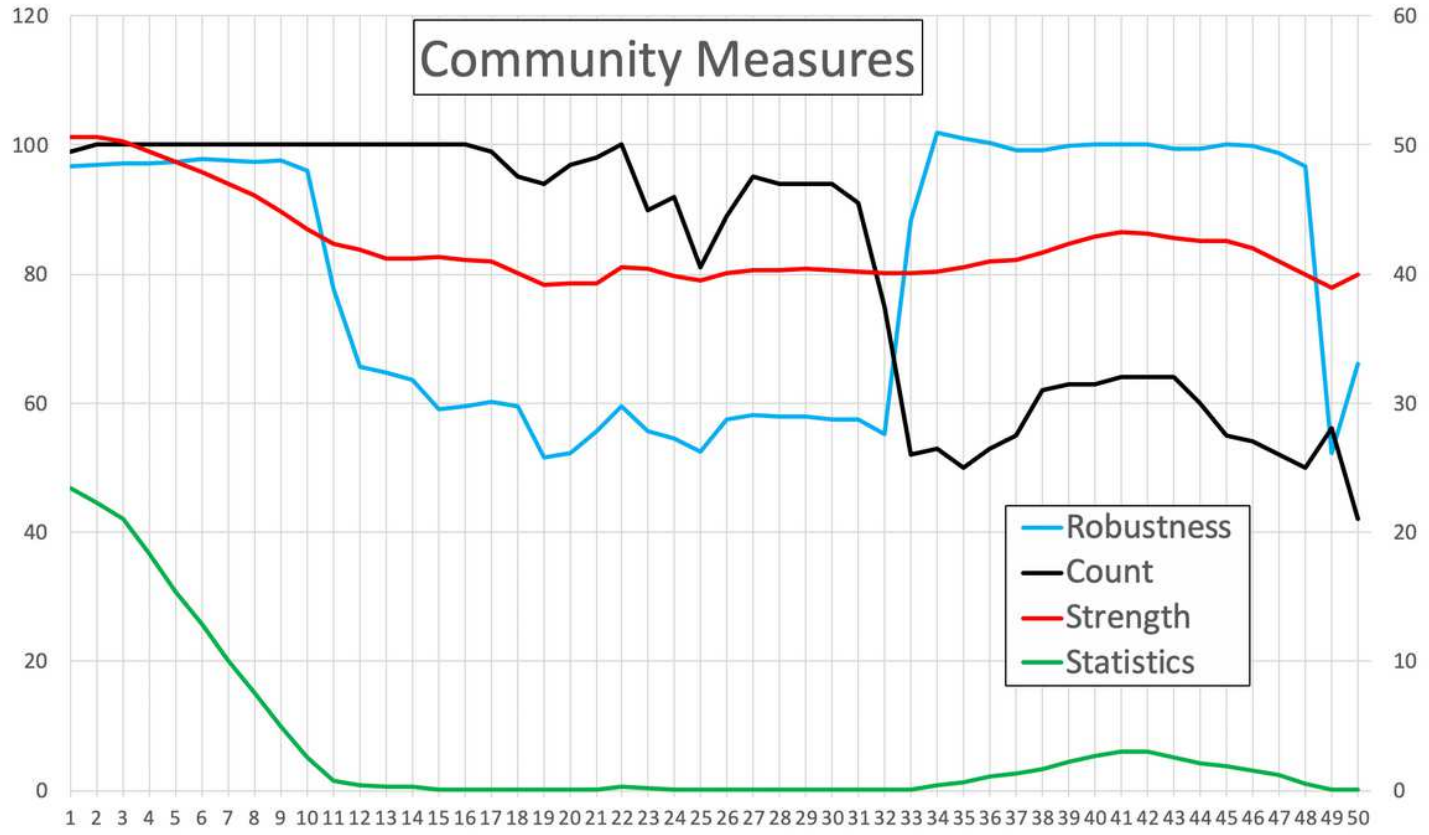


Figure 11

Values of the community measures for the Les Misérables network. Compositions of the divisions are listed in Table 5. Fifty highly ranked divisions of the 201 division detected in 100 computer runs with uniformly U(0.05,0.01) distributed random link weights are shown in the figure.

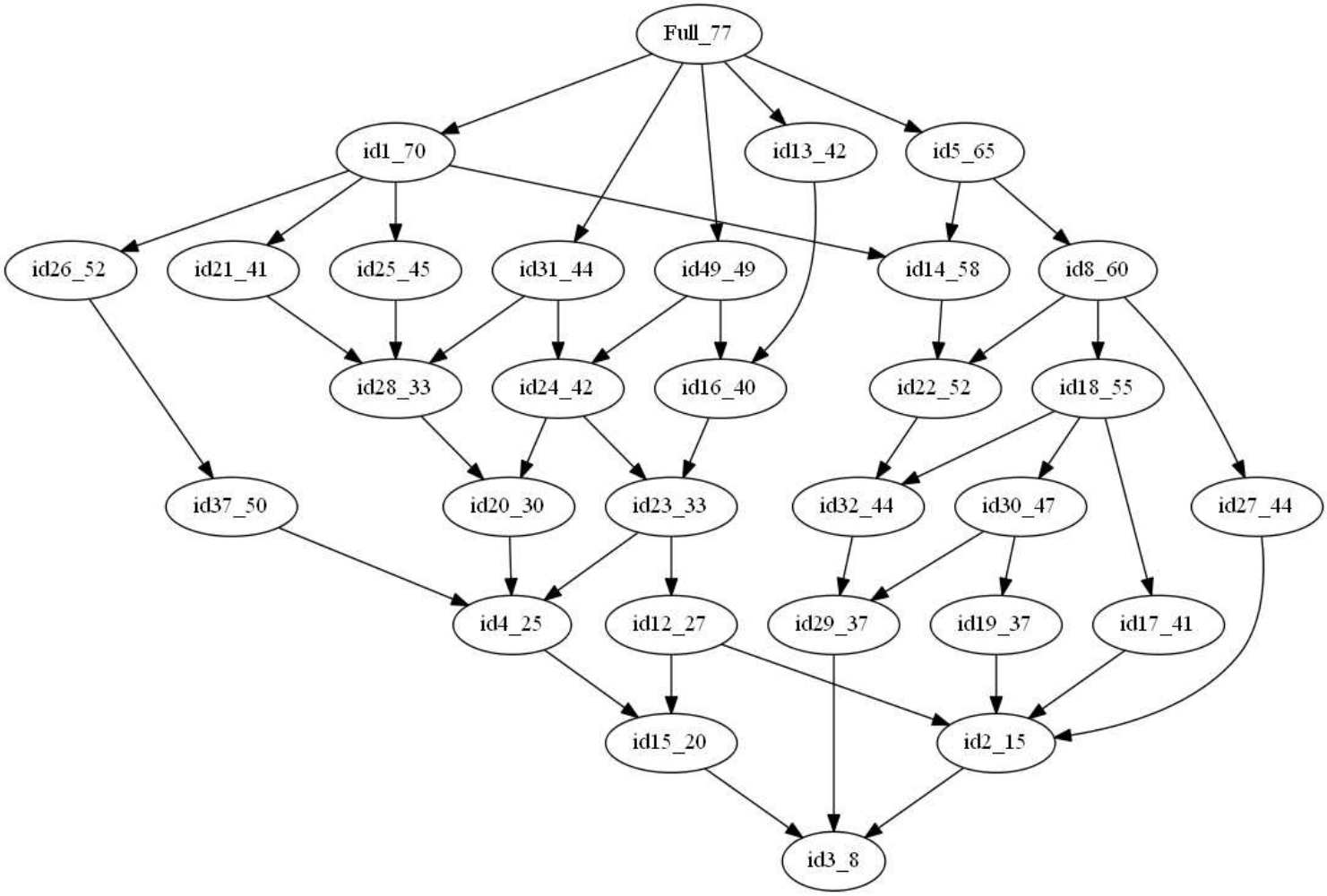


Figure 12

Hierarchical structure of 29 sub-communities of the Les Misérables network (only half of the hierarchical structure is shown). Divisions indicated by id-numbers are listed in Table 5 with bold ids. Results are for the influence spreading model with link weights 0.05. Sizes of sub-communities are also displayed in the graph.

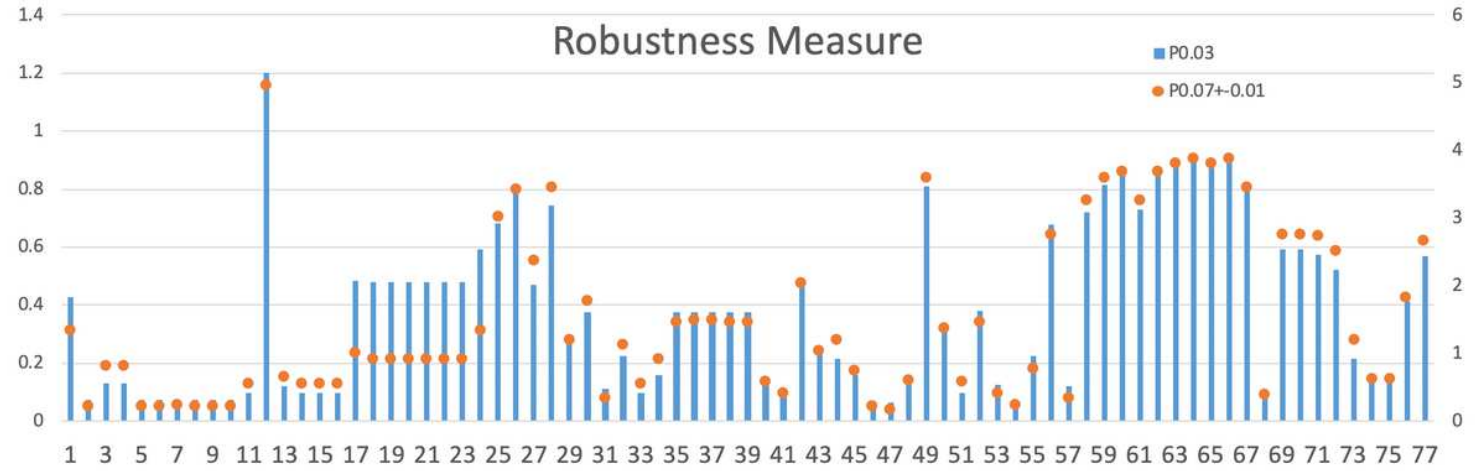


Figure 13

. Robustness values of nodes in the Les Misérables network from the influence spreading model with link weights 0.03 and uniformly distributed random link weights between [0.06, 0.08].

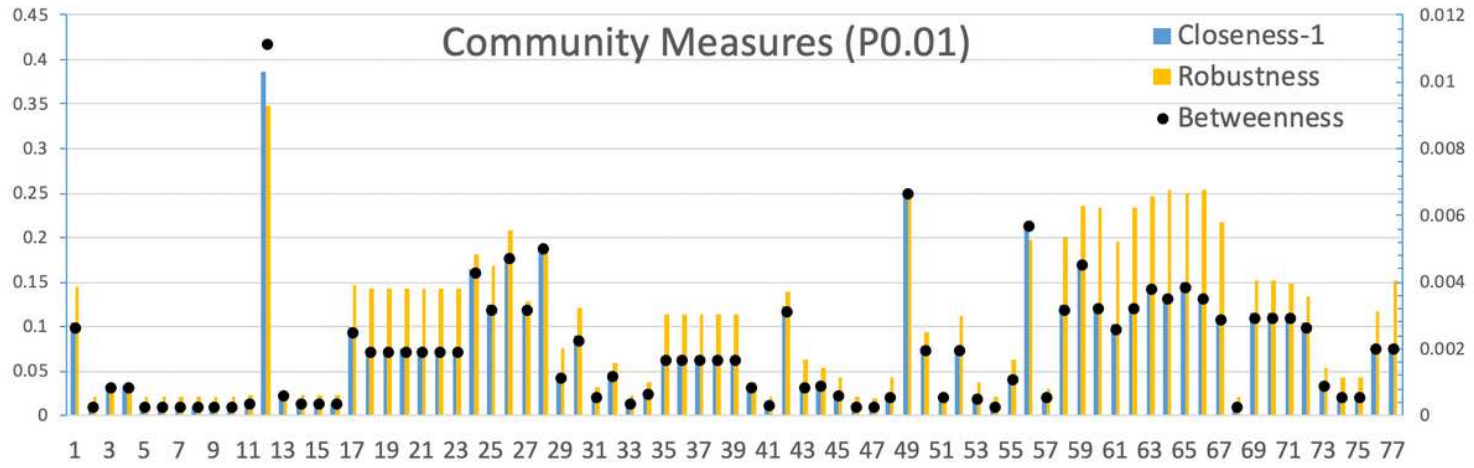


Figure 14

Closeness and betweenness centrality measures compared with the robustness measure values of nodes in the Les Misérables network. The influence spreading model is used with link weights 0.01. Detailed composition of robustness values are shown in Fig. 15.

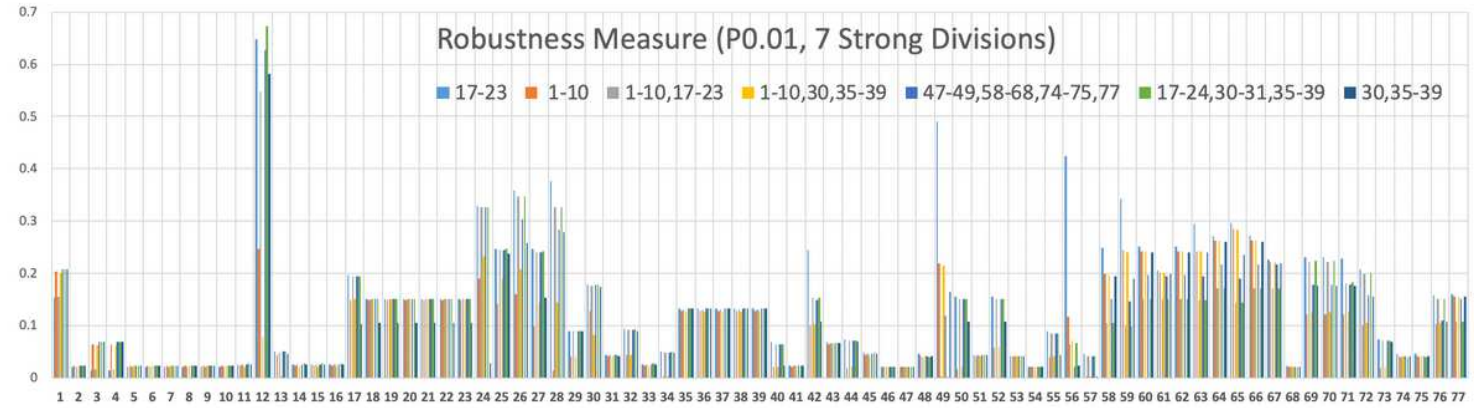


Figure 15

Robustness values of nodes for seven strong divisions in the Les Misérables network. Corresponding average values of robustness are shown in Fig. 14.

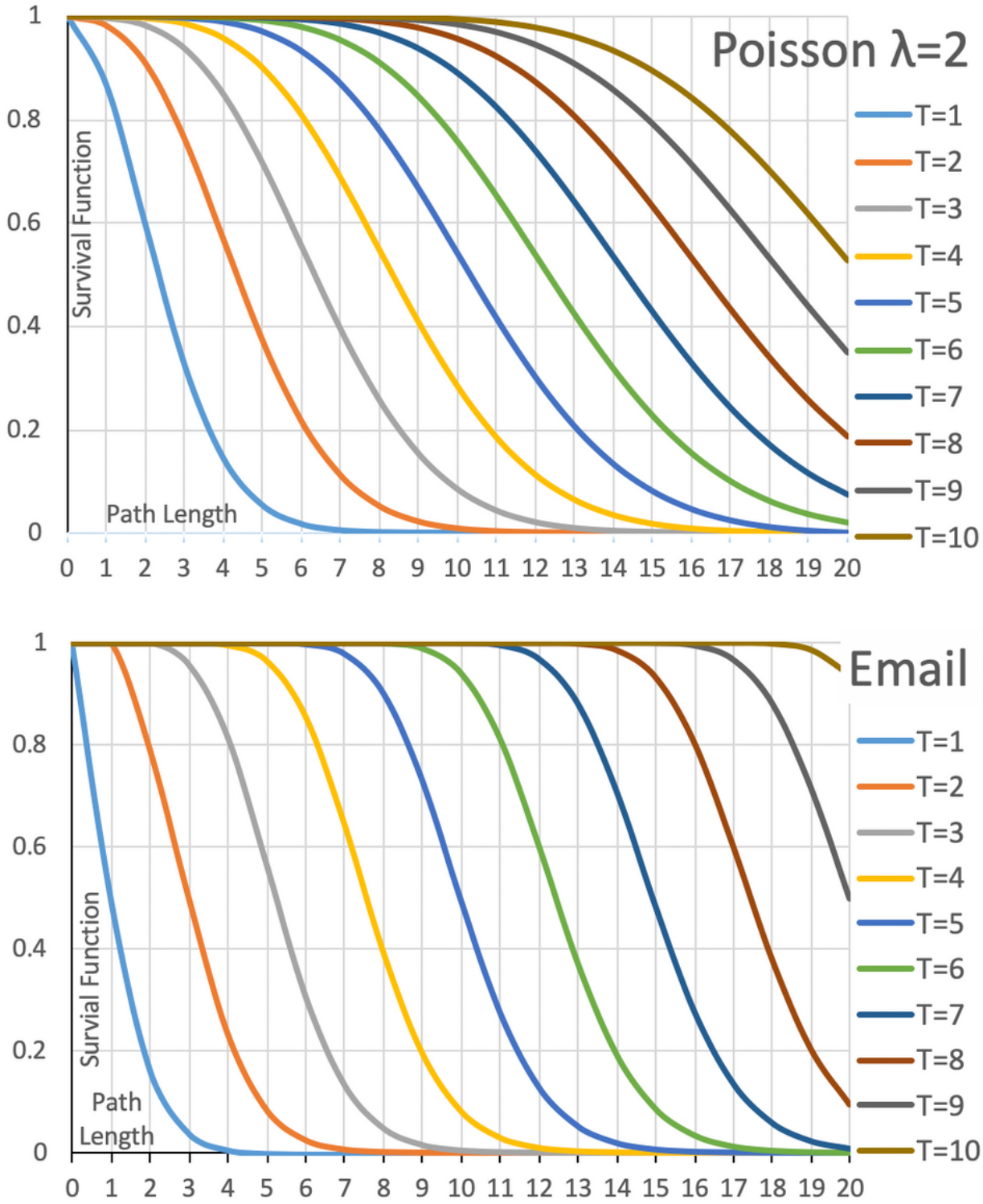


Figure 16

Survival distribution functions for the Poisson and e-mail forwarding distribution functions as a function of path length for time values T=1,…,10. The event rate parameter value of λ=2 for the Poisson distribution has been used to get comparable results with these path lengths.

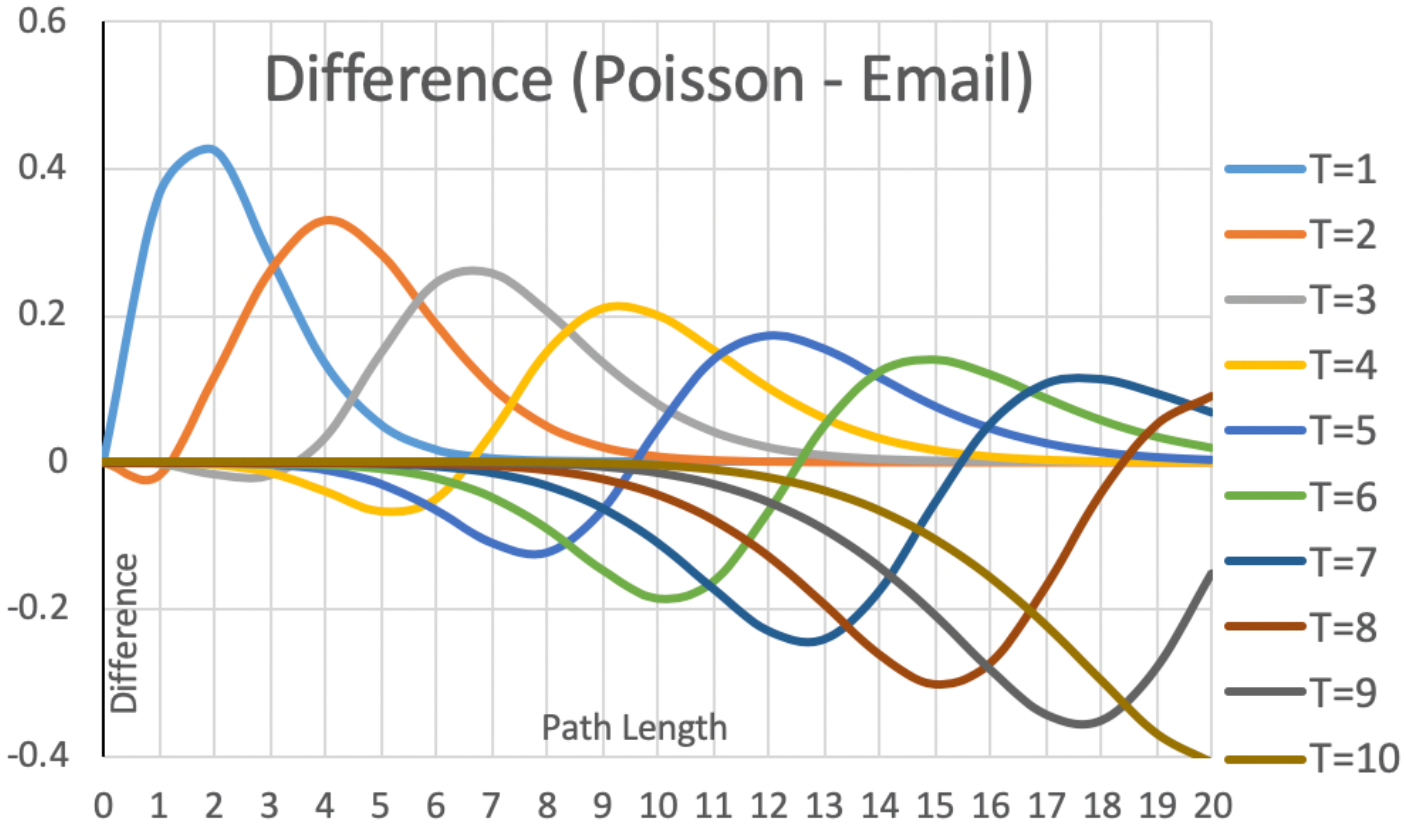


Figure 17

Difference between Poisson and E-mail forwarding survival distribution functions as a function of path length for time values T=1,…,10.

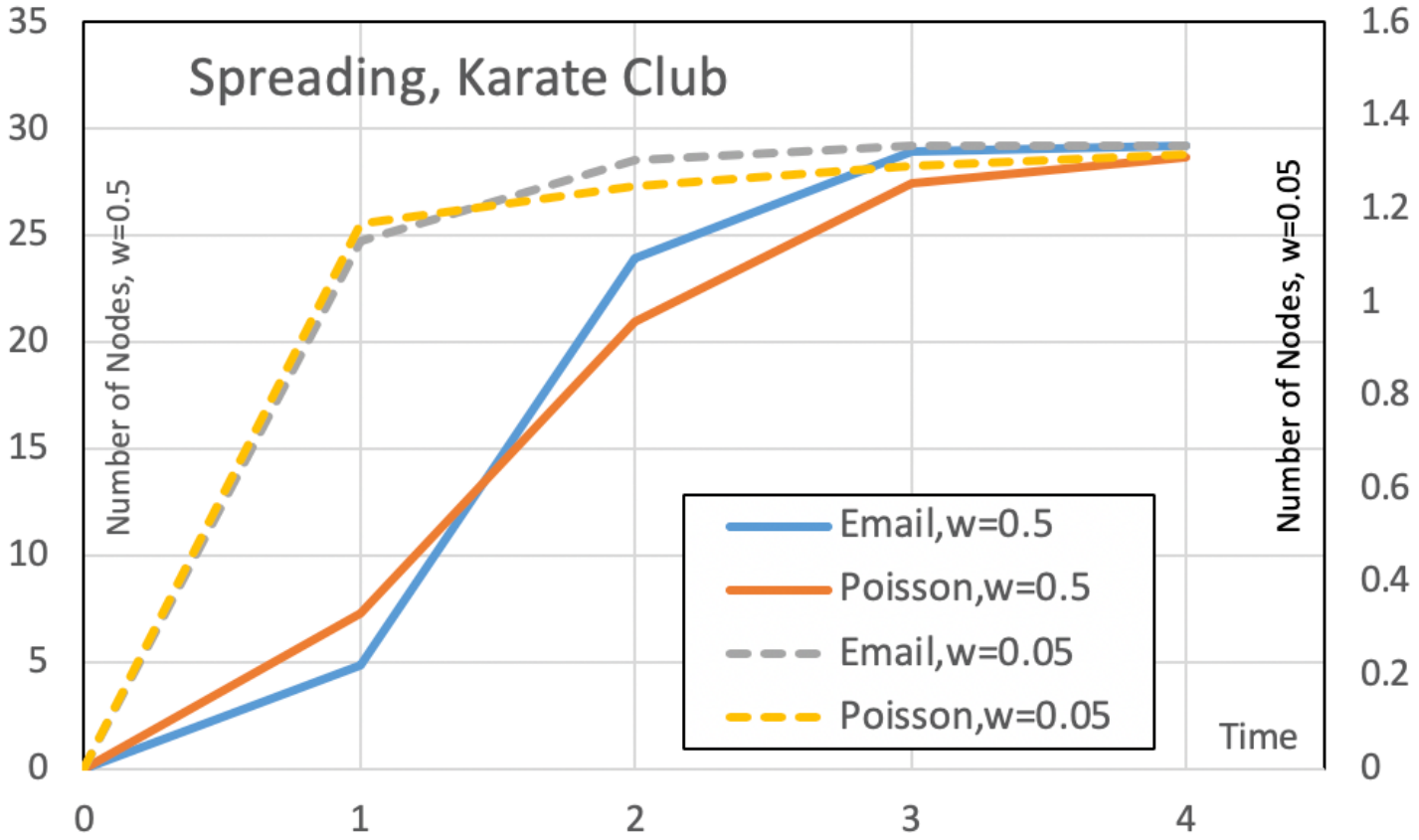


Figure 18

Spreading in the Zachary's karate club network for the e-mail forwarding process and the Poisson process. Y-axis on the left (right) shows the expected number (normalised) of influenced nodes for w=0.5 (w=0.05).

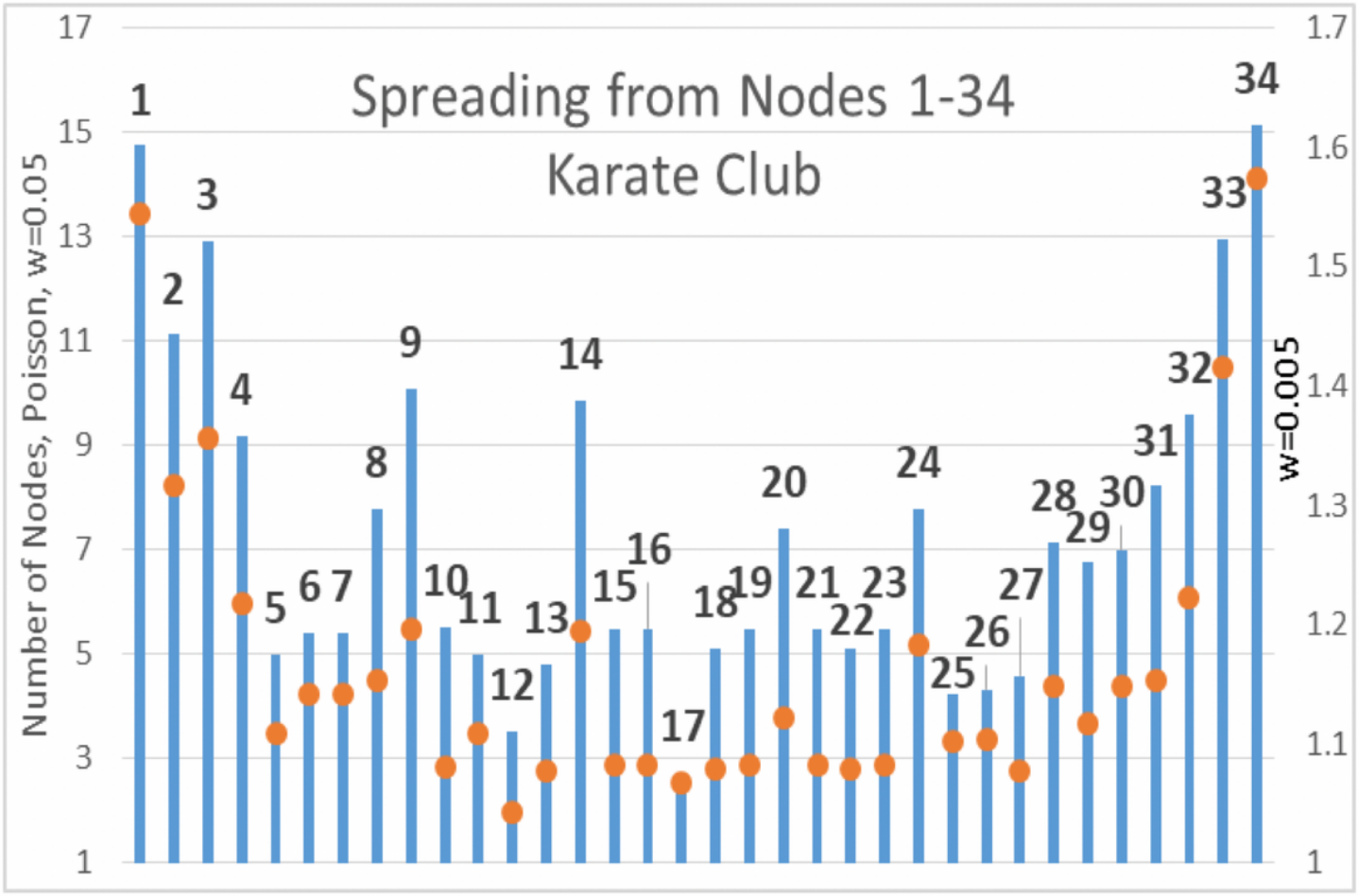


Figure 19

Spreading from individual nodes 1-34 to other nodes (including the node itself) for the equilibrium state when time approaches infinity. Expected number (normalised) of influenced nodes for w=0.05 on the left Y-axis and for w=0.005 on the right. Numerical values for w=0.05 are shown as bars and for w=0.005 as circles.

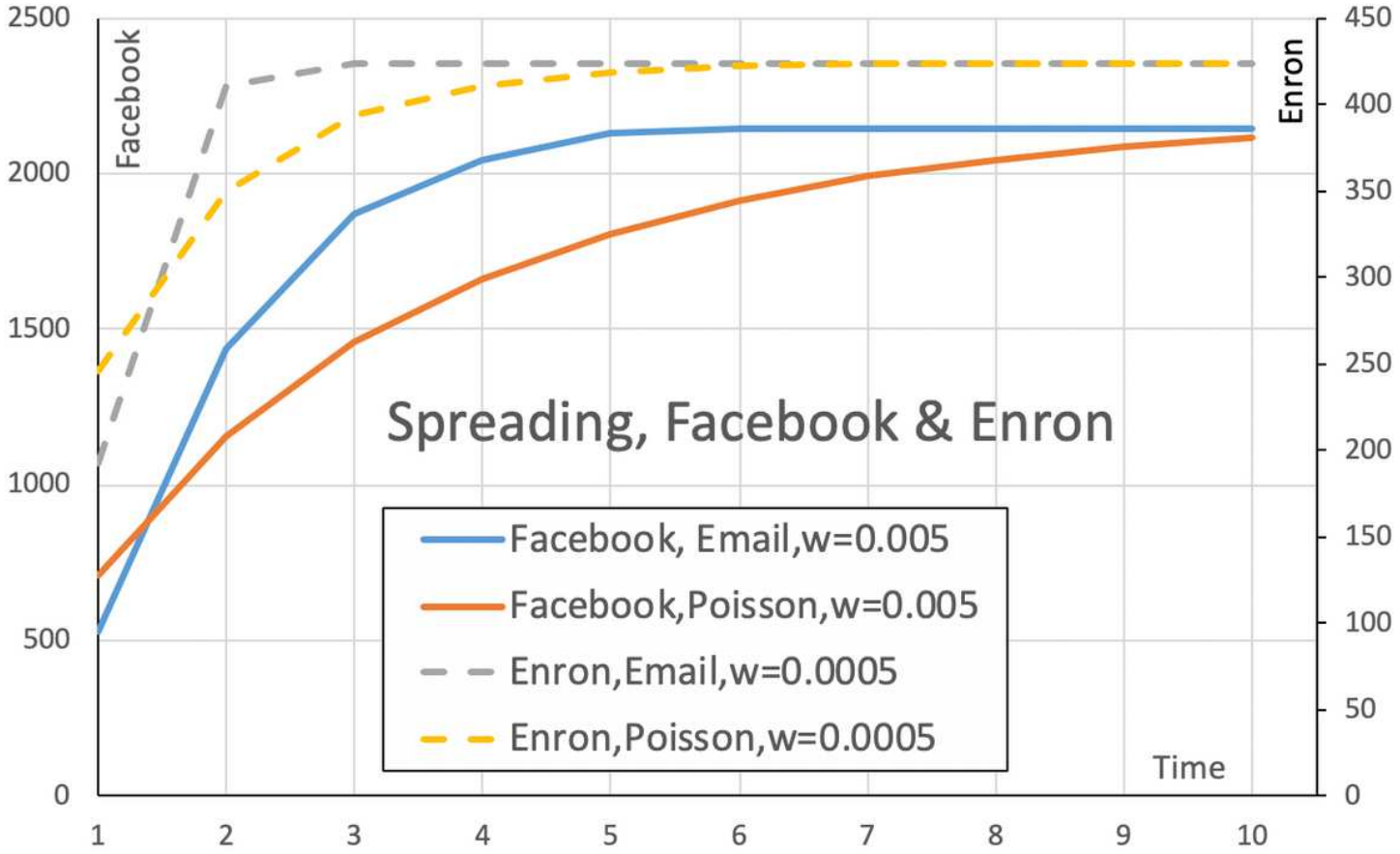


Figure 20

Spreading in the Facebook and Enron networks for the Poisson process and the e-mail forwarding process. Y-axis on the left (right) shows the expected number (un-normalised) of influenced nodes for the Facebook network (Enron network).