



OPEN

Interplay between tie strength and neighbourhood topology in complex networks

Maciej J. Mrowinski[✉], Kamil P. Orzechowski, Agata Fronczak & Piotr Fronczak

Granovetter's weak ties theory is a very important sociological theory according to which a correlation between edge weight and the network's topology should exist. More specifically, the neighbourhood overlap of two nodes connected by an edge should be positively correlated with edge weight (tie strength). However, some real social networks exhibit a negative correlation—the most prominent example is the scientific collaboration network, for which overlap decreases with edge weight. It has been demonstrated that the aforementioned inconsistency with Granovetter's theory can be alleviated in the scientific collaboration network through the use of asymmetric measures. In this paper, we explain that while asymmetric measures are often necessary to describe complex networks and to confirm Granovetter's theory, their interpretation is not simple, and there are pitfalls that one must be wary of. The definitions of asymmetric weights and overlaps introduce structural correlations that must be filtered out. We show that correlation profiles can be used to overcome this problem. Using this technique, not only do we confirm Granovetter's theory in various real and artificial social networks, but we also show that Granovetter-like weight-topology correlations are present in other complex networks (e.g. metabolic and neural networks). Our results suggest that Granovetter's theory is a sociological manifestation of more general principles governing various types of complex networks.

While this is not always the case, the weights of edges in networks are usually quantitative expressions of the mutual relationship between nodes. Be it the number of scientific collaborations between authors or the number of mentions in a social network, the weight of an edge often signifies the strength of the connection between nodes. It stands to reason that this strength must, in some way, correlate with the network's structure—specifically, with the relative position of nodes and their neighbourhoods within the network.

Mark Granovetter, in his famous work strength of weak ties^{1,2}, introduced a theory which aims to explain the aforementioned link between the weights of edges and the topology of the network. An example that illustrates Granovetter's hypothesis can be found in Fig. 1, which shows two fully connected clusters of nodes. According to Granovetter, since virtually all nodes in each cluster have the same neighbourhoods, we should expect that edge weights (tie strengths) within clusters will be high. A single edge also connects the clusters. Edge weight of this connection should be low, as nodes at both sides of the link do not share any neighbours. Granovetter's theory also states that weak ties, like the one connecting the clusters in our example, are crucial to the diffusion of information in the network and nodes that have access to such ties have an advantage over those that do not. In this work, however, we are only interested in the first part of the theory, that is in weight-topology correlations.

In more formal terms, the first part of Granovetter's theory states that edge weight is positively correlated with the overlap of the neighbourhood of two connected nodes (see Fig. 2a). The overlap between neighbourhoods of node i and node j is defined in the following way³

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_j - 1) - n_{ij}}, \quad (1)$$

where n_{ij} is the number of common neighbours of nodes i and j , k_i and k_j are degrees of nodes i and j . It is worth noting that overlap, as defined above, is a symmetric measure

$$O_{ij} = O_{ji}, \quad (2)$$

Faculty of Physics, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland. ✉email: maciej.mrowinski@pw.edu.pl

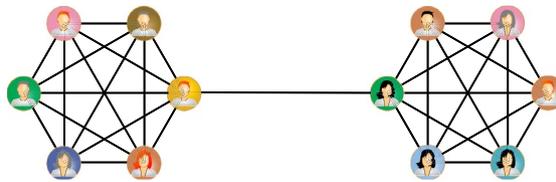


Figure 1. An example illustrating Granovetter’s theory. A detailed description can be found in the text.

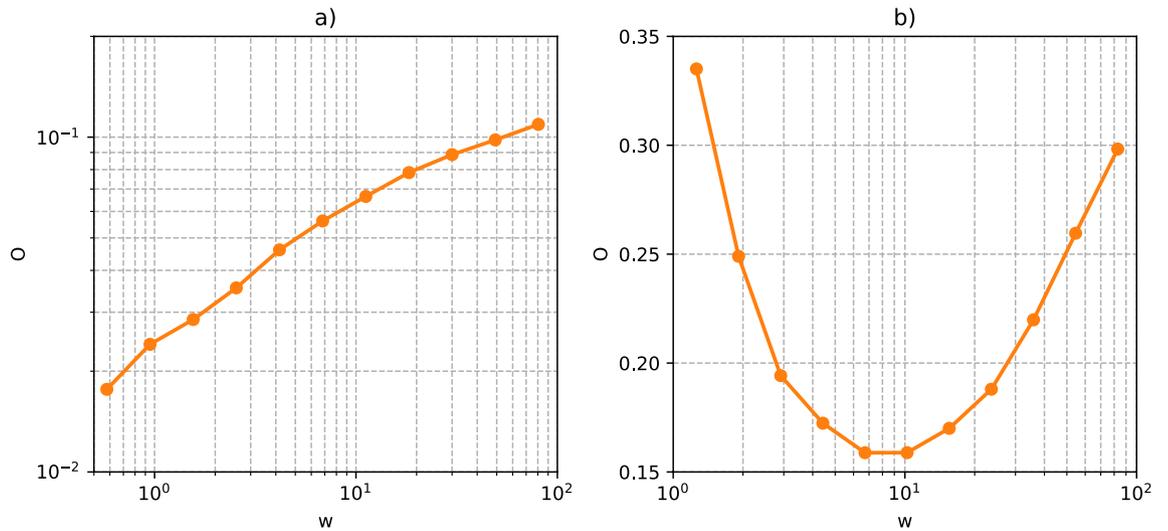


Figure 2. The average value of symmetric overlap as a function of symmetric weight for (a) Twitter, in which Granovetter’s theory holds, and (b) DBLP. In the second panel, for the majority of samples (cf. heatmap in Fig. 7a), the symmetric overlap decreases with the symmetric weight, defying Granovetter’s theory. Note that the white curve in Fig. 7a is equivalent to the curve in panel (b) in this figure (the only difference is that in this figure, the Y axis is not logarithmic, and we averaged the results arithmetically, instead of using the geometric mean like in Fig. 7. The curve in panel (a) is the same as the white curve in Fig. 5a.

and we will refer to it as symmetric overlap to emphasize this fact. Similarly, weights w are also assumed to be symmetric, that is

$$w_{ij} = w_{ji}. \tag{3}$$

Granovetter’s theory in this form—that is, a monotonically increasing relation between O_{ij} and w_{ij} —has been empirically confirmed, to various extents, in real social networks^{3–10}, like the mobile communication network. However, there are also counterexamples to the theory, one of which is the scientific collaboration network^{11–13}. In this network, nodes represent authors, and an edge connects two authors if they co-authored at least one manuscript. The symmetric weight w_{ij} equals the number of manuscripts co-authored by authors i and j .

At first glance, the scientific collaboration network seems to defy Granovetter’s theory, as neighbourhood overlap, on average, decreases with edge weight for the majority of edges (see Fig. 2b). We have shown, however, that this supposed disagreement stems from improper definitions of weights and overlaps¹⁴. Or, more specifically, from the fact that symmetric measures cannot properly describe the properties of this network.

Figure 3 illustrates the problem with symmetric measures. Panel (a) shows two nodes, left l and right r , with degrees $k_l = k_r = 4$. The nodes share two common neighbours. In this case, symmetric overlap equals

$$O_{lr} = O_{rl} = \frac{2}{(4 - 1) + (4 - 1) - 2} = \frac{2}{4} = \frac{1}{2}. \tag{4}$$

Since the degrees of both nodes are identical, the two common neighbours constitute the same fraction of the neighbourhood of each node. In such a scenario, symmetric overlap accurately reflects this observation from the standpoint of both nodes. However, symmetric overlap fails when assessing nodes with vastly different degrees. In Fig. 3b, the left node with $k_l = 4$ shares two common neighbours with the right node, whose degree is $k_r = 16$. Intuitively, the left node should assign much greater significance to the two common neighbours than the right node. Symmetric overlap cannot be used as a measure of this significance, as for both nodes it equals

$$O_{lr} = O_{rl} = \frac{2}{(4 - 1) + (16 - 1) - 2} = \frac{2}{16} = \frac{1}{8}. \tag{5}$$

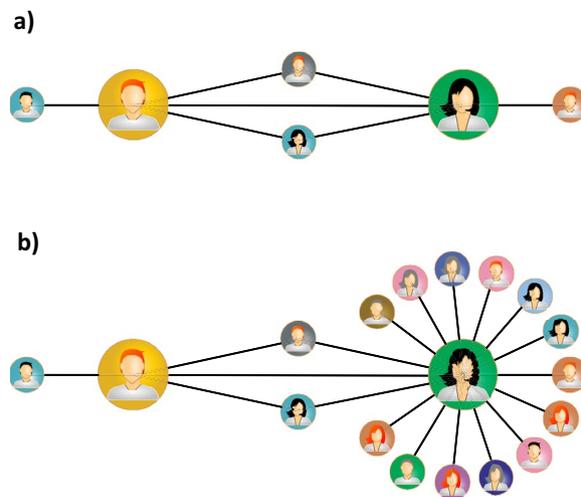


Figure 3. Examples of two connected nodes with (a) similar and (b) vastly different degrees.

It is a low value, clearly skewed towards the node with the higher degree.

These two examples show that symmetric measures work properly when dealing with homogeneous networks, where we compare similar nodes—as was the case for many networks in which Granovetter’s theory was proven to hold. Non-homogeneous networks, like the scientific collaboration network, which is scale-free^{15,16} (there are often nodes with highly different degrees on both sides of an edge), require a different approach. The innate asymmetry of these networks suggests that one must use asymmetric measures instead of symmetric ones. In¹⁴, we introduced the asymmetric overlap Q :

$$Q_{ij} = \frac{n_{ij}}{k_i - 1}, \quad (6)$$

with

$$Q_{ij} \neq Q_{ji}. \quad (7)$$

Returning to the example from Fig. 3b, asymmetric overlap for the left node is

$$Q_{lr} = \frac{2}{4 - 1} = \frac{2}{3}, \quad (8)$$

while for the right node, we have

$$Q_{rl} = \frac{2}{16 - 1} = \frac{2}{15}. \quad (9)$$

These two values of overlap properly convey the importance of the shared neighbourhood from the perspective of each node separately. Asymmetric overlap reflects the asymmetric relationships of authors in the scientific collaboration network (and other non-homogeneous networks). What can be a large fraction of collaborators from the perspective of one author, can be a negligible fraction from the perspective of another author.

Symmetric definitions of weights suffer from similar issues in non-homogeneous networks. Symmetric weight in the scientific collaboration network usually equals the number of collaborations (co-authored articles) between two authors. However, the importance of a single collaboration depends on the total number of collaborations. If someone wrote only one paper in collaboration with an author who published tens or hundreds of manuscripts, then intuitively, the strength of that tie (the weight of the edge) should be greater from the perspective of the former author. Thus, in¹⁴, we also introduced the asymmetric definition of weight v :

$$v_{ij} = \frac{w_{ij}}{m_i}, \quad (10)$$

where w_{ij} is the symmetric weight, and m_i is the number of papers published by the i -th author. For asymmetric weights, we also have

$$v_{ij} \neq v_{ji}. \quad (11)$$

Using asymmetric overlaps and asymmetric weights, we showed that Granovetter’s theory holds in the scientific collaboration network. We also postulated that these are natural and intuitive tools capable of properly describing scale-free networks, with application to other problems, eg. link prediction¹⁷. However, these asymmetric definitions introduce a certain non-obvious issue, especially when it comes to confirming Granovetter’s theory.

The nature of this theory—or rather, the nature of the correlation between weights and overlaps postulated by Granovetter—is sociological. That is, the correlation must stem from actual social interactions between entities represented by nodes in the network. In contrast to that, measures defined in (6) and (10) introduce structural correlations to the mix—correlations that result from the topology of the network. It is not unreasonable to assume that the number of papers published by an author (m_i) will relate in some way to the total number of collaborators (k_i)—intuitively, one can expect a positive correlation between these two variables. It raises the following questions: What are we really observing if we detect a correlation between asymmetric overlap and asymmetric weight? What is the source of that correlation? Are we truly confirming Granovetter’s theory, or are we merely misinterpreting the effects of the network’s topology? This paper aims to dispel these doubts using tools introduced in the next section.

Methods

Let us reiterate the problem mentioned in the Introduction and define it in a clearer and more tangible way. The main assumption behind Granovetter’s theory is that weights in social networks are not assigned to edges randomly. Instead, they quantify the strength of interpersonal interactions and follow various patterns dictated by the nature of these interactions. One such pattern is that the strength of interactions should be directly tied to the overlap between social circles of nodes. The higher the overlap, the greater the strength of interaction. It is an intuitive and relatable conclusion—for example, ties within a family, which is a densely connected social circle, should be stronger than ties within a workplace.

Assuming that Granovetter’s theory is correct, we could expect that in a network in which weights are assigned completely at random, the correlation between overlap and weight does not exist at all. By the same token, if we were to randomize weights in a network by shuffling them among the edges, such a procedure should also destroy the correlation between overlaps and weights. Unfortunately, while this is true for symmetric measures, asymmetric weights and overlaps still exhibit correlation even with randomised weights. The source of these correlations was mentioned in the previous section, and it is the structural correlation between m_i and k_i , which are in the denominators of the asymmetric measures.

In fact, for reasons that will be explained in detail in the next section, we will use a definition of asymmetric weight different to the one introduced in¹⁴. In this work, asymmetric weight will be defined as (cf. Eq. 10)

$$v_{ij} = \frac{w_{ij}}{s_i}, \quad (12)$$

where s_i is the strength of the i -th node

$$s_i = \sum_j w_{ij}. \quad (13)$$

Here, the structural correlations are even clearer. Since $s_i \propto k_i$ ¹⁸ (for example, if we assume that all weights $w_{ij} = 1$, then s_i is equal to k_i),

$$v_{ij} \propto \frac{1}{s_i}, \quad (14)$$

and

$$Q_{ij} \propto \frac{1}{k_i}, \quad (15)$$

we must have

$$Q_{ij} \propto v_{ij}. \quad (16)$$

The existence of the correlation between Q_{ij} and v_{ij} is largely independent of the distribution of symmetric weights w_{ij} in the graph. That is, if we were to shuffle existing weights between edges or assign completely new weights to edges according to some probability distribution, this structural correlation would still be present.

The challenge is, then, to decouple the structural correlations from Granovetter-like social correlations while keeping the asymmetric definitions of strengths and overlaps. Thankfully, this is hardly a new kind of problem, and there are tools capable of dealing with similar issues. More specifically, we are going to employ so-called correlation profiles^{19,20}, which were used before to study mixing patterns in complex networks (correlations between node degrees at the ends of the same edge)^{21–24}.

The idea behind correlation profiles is simple but powerful. One needs to compare the properties of the actual network with the properties of its randomised realisations (the null model). If the null model is chosen correctly, then the difference between random realisations and the actual network should result not from structural correlations but, in our case, from sociological processes (which are not present in the null model) that govern the assignment of weights to edges.

Correlation profiles are constructed using two simple ratios. If we want to study some pattern p observed in a network, then we have to compare the number $N(p)$ of occurrences of that pattern in the actual network with the average number $\langle N_r(p) \rangle$ of occurrences of the same pattern in randomised realisations of the network. Using these two numbers, we can define the ratio

$$R(p) = \frac{N(p)}{\langle N_r(p) \rangle}. \quad (17)$$

If $R(p)$ is close to 1, then there is no significant difference between the null model and the actual network. It follows that pattern p is associated with properties captured in the null model. On the other hand, if $R(p)$ is higher or lower than 1, then there are mechanisms in the actual network that are responsible for the creation (or dissolution) of pattern p that are not present in the null model.

The second ratio—Z-score—is defined as

$$Z(p) = \frac{N(p) - \langle N_r(p) \rangle}{\Delta N_r(p)}, \quad (18)$$

where $\Delta N_r(p)$ is the standard deviation of $N_r(p)$ in the randomised realisations of the network. This ratio determines the statistical significance of $R(p)$.

In most cases, correlation profiles are represented as two-dimensional images. To give a more concrete example, if we want to study the relation between overlap Q and weight v , we divide the $Q - v$ plane into two-dimensional bins of equal size on a log-log scale (we use a logarithmic scale because Q and v values span multiple decades). Patterns p correspond to pairs (v, Q) (each edge in the network introduces two such pairs) falling into corresponding bins.

We count the number of points $N(p_i)$ that fall into the i -th bin in the actual network (here, p_i denotes a pattern corresponding to a point falling into the i -th bin). Next, we create many random realisations of the network by shuffling symmetric weights and average over these realisations the number of points $N_r(p_i)$ that fall into the corresponding bin. Dividing these two numbers gives us the ratio $R(p_i)$ for the i -th bin. We repeat this procedure for each bin (using the same random realisations), which gives us the full correlation profile. Z-scores are calculated in the same way.

An illustration of $R(p_i)$ calculation can be found in Fig. 4. In this example, we concentrate on the middle bin. When weights are shuffled among edges during the creation of randomised graph instances (the null model), points on the $Q - v$ diagrams change their positions. However, they only move along the v axis. The overlaps, which are independent of weights, do not change. Since the network's topology is fully retained during weight shuffling, the null model leaves the structural correlations intact. In Fig. 4, points that will move into the middle bin after shuffling are orange, while the point that will move out of the middle bin is green. Arrows indicate where each of the relevant points will end up after shuffling. Thus, for the middle bin, we have

$$R(\text{middle bin}) = \frac{1}{4}. \quad (19)$$

This value of R suggests that the processes responsible for the distribution of weights in the actual network remove points from the middle bin when compared with a random instance of the network, possibly prioritizing other bins in the diagram. While it is an oversimplification (especially since we used only one randomised network instance instead of an entire ensemble, as required by the definition in Eq. (17)), this example demonstrates the main idea behind the correlation diagrams. The non-structural correlations can be singled out by comparing the positions of points on the $Q - v$ diagrams corresponding to the actual network and its randomised instances.

Datasets

In¹⁴, we studied the validity of Granovetter's theory only in the scientific collaboration network. In this work, wanting to test both the theory itself and the applicability of correlation profiles on a variety of different networks, we used 8 datasets in total:

- Twitter (source: <https://doi.org/10.6084/m9.figshare.13308428.v1>)—the network of Twitter mentions¹². Nodes represent Twitter users and weights correspond to the number of mentions.
- DBLP (source: <https://www.aminer.org/citation>)—the scientific collaboration network (version 12). It contains metadata about scientific articles²⁵, including lists of authors and references. Nodes represent authors;

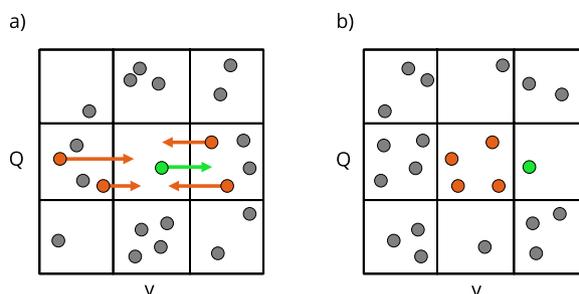


Figure 4. An example illustrating the creation of correlation profiles—points in bins (a) before weight shuffling and (b) after shuffling.

two authors are connected if they co-authored at least one paper. Symmetric weight is equal to the total number of papers co-authored by two authors.

- Actor Movies (source: <http://konect.cc/networks/actor-movie/>)—nodes represent actors, two actors are connected if they appeared in the same film. Symmetric weight is equal to the number of films in which actors worked together.
- Record Labels (source: <http://konect.cc/networks/dbpedia-recordlabel/>)—nodes represent music artists, two artists are connected if they performed under the same record labels. Symmetric weight is equal to the number of record labels under which artists worked together.
- The Marvel Universe Social Network (source: <https://www.kaggle.com/datasets/csanhueza/the-marvel-universe-social-network/>)—nodes represent heroes, two heroes are connected if they appeared in the same comic²⁶. Symmetric weight is equal to the number of comics in which heroes appeared together.
- Flights (source: commercially available at <https://www.icao.int/>)—network of passenger flights. Nodes represent airports, and weights correspond to the volume of traffic (number of passengers) between airports.
- Metabolic Network (source: <https://www.ebi.ac.uk/biomodels/MODEL6399676120>)—where nodes represent reactants, connected by an edge when they take part in the same reaction²⁷. Symmetric weight equals the number of reactions sharing two given reactants.
- *Caenorhabditis elegans* (source: <http://konect.cc/networks/dimacs10-celegansneural/>)—the neural network of *Caenorhabditis elegans*²⁸. Nodes represent neurons, and an edge links two neurons if a synapse or gap junction connects them. Weights correspond to the total number of connections between neurons.

Not all of these networks are social networks, and some are artificial social networks. However, they all exhibit a Granovetter-like relationship between overlaps and weights. Table 1 contains information about the sizes of the largest connected components in the networks. We restricted our analysis to the largest components because it is likely that smaller components (especially in networks like DBLP) stem from the incompleteness of data and may introduce unwanted noise and artefacts to the results. However, correlation diagrams for entire networks are qualitatively equivalent to the ones presented in the manuscript.

Some of the networks we used can be represented as bipartite graphs (e.g. DBLP, Actor Movies—virtually all collaboration networks can be stored in this form) and recovered via appropriate projections^{29,30}. These networks are undirected and have a well-defined notion of symmetric weight. One can also easily use (10) to define asymmetric weights in such networks, with m_i equal to the degree of node i in the bipartite representation of a graph (which corresponds to the total number of collaborations for a given node—e.g. movies or scientific manuscripts). On the other hand, networks like Twitter or Flights are inherently directed, cannot be expressed as bipartite graphs and, consequently, Eq. (10) cannot be applied in a meaningful way.

In order to standardise our approach to the networks under study and overcome problems associated with Eq. (10), we decided to symmetrise all directed networks and assumed that symmetric weight in their undirected equivalent is equal to the average of weights in both directions:

$$w_{ij} = \frac{V_{ij} + V_{ji}}{2}, \quad (20)$$

where V_{ij} and V_{ji} are weights of directed edges. At the same time, we abandoned the definition of asymmetric weight introduced in¹⁴, and settled on definition (12) instead (where asymmetry is achieved by normalising symmetric weight—that is by dividing it by the strength of a node). While it may seem as counter intuitive—directed networks are converted to undirected ones using Eq. (20), only to be converted again to directed networks using Eq. (12)—this approach allows us to treat all networks, both directed and undirected ones, in the same way and to compare results.

Results

Correlational profiles for Twitter, a real social network, are shown in Fig. 5. Panels (a) and (b) contain, for comparison with their asymmetric counterparts, heatmaps of the symmetric overlap O as a function of symmetric weight w for the actual network and the null model. It is worth noting that, in many cases, symmetric weights are integers, and edges are often characterized by the same weight values. This makes edges indistinguishable

Dataset	Nodes	Edges	Type
Twitter	510,136	5,210,278	Directed
DBLP	2,851,120	24,965,776	Bipartite
Actor movies	374,511	30,029,678	Bipartite
Record labels	11,078	117,798	Bipartite
Marvel	6403	343176	bipartite
Flights	1292	24,925	Directed
Metabolic network	3160	29,210	Bipartite
<i>Caenorhabditis elegans</i>	297	4296	Undirected

Table 1. Sizes of largest connected components in the datasets.

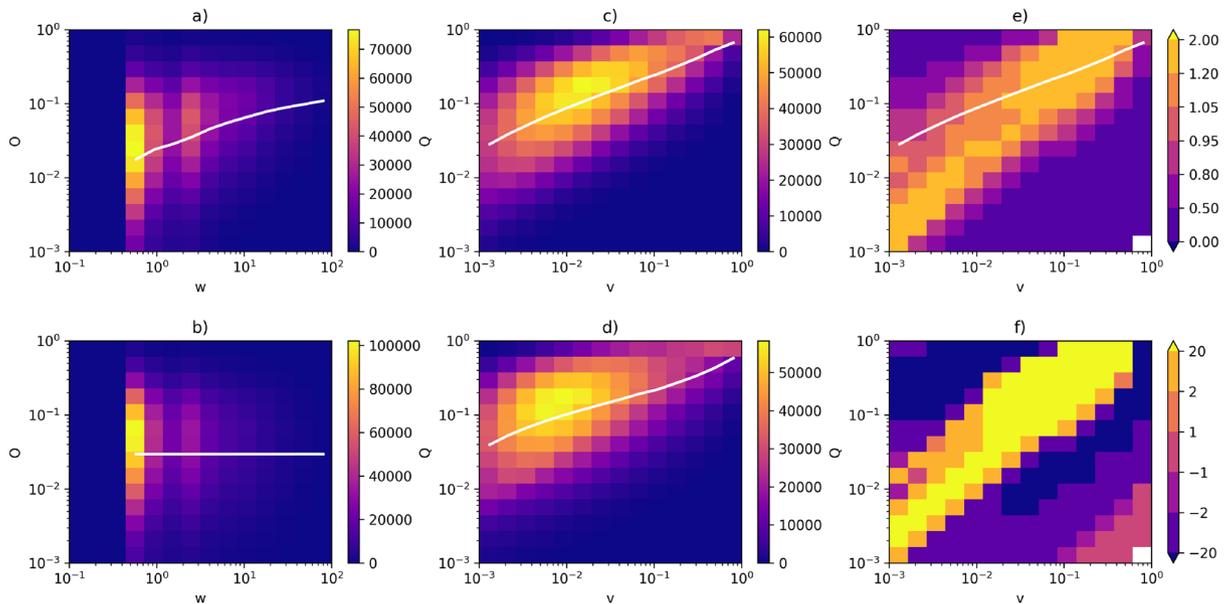


Figure 5. Correlation profiles for Twitter. **(a)** Heatmap for the actual network—symmetric weights. **(b)** Heatmap for the null model (randomised network)—symmetric weights. **(c)** Heatmap for the actual network— asymmetric weights. **(d)** Heatmap for the null model (randomised network)—asymmetric weights. **(e)** Correlation profile (R). **(f)** Z-score (Z). The white lines in **(a,b)** correspond to the average O as a function of w , on **(c,d)**—the average Q as a function of v . The line in panel **(e)** is the same as in **(c)**.

from one another, which is a problem associated with using symmetric weights. Asymmetric weights are free of this issue, which is their additional benefit.

Panel **(c)** contains heatmaps for the asymmetric overlap Q as a function of asymmetric weight v . A clear, Granovetter-like relation is visible—overlap increases with weight. However, almost the same relation is present in panel **(d)**, which contains the equivalent heatmap for the null model (the same network with shuffled edges). These two panels show the root of the issue with the asymmetric definitions of weights and overlaps. Granovetter's theory dictates how weights should be distributed in a graph. If the theory is correct, then we should reasonably expect that there is no correlation between Q and v in the null model—the shuffling procedure should destroy any deliberate (from the perspective of Granovetter's theory) placement of weights. Unfortunately, such a correlation is also present in the null model due to the network's topology. Moreover, at first glance, the relation between Q and v seems to be very similar in the actual network and the null model.

This is where the correlation profiles come into play. By comparing panels **(c)** and **(d)**—that is, by dividing counts in bins in **(c)** by counts in corresponding bins in **(d)**, which creates the correlation profile R , Eq. (17)—we can easily find the differences between the null model and the real network. Panel **(e)** shows such a profile. We can also see a Granovetter-like relation visible there—linear (on a log-log scale) clusters of bins such that more edges fall into these clusters in the actual networks than in the null model. It strongly suggests that Granovetter's theory is indeed correct and that sociological processes that govern the distribution of weights in real networks result in higher weights assigned to edges with higher values of overlaps. These results are statistically significant, which is confirmed by Z-scores in panels **(f)**.

We calculated correlation profiles and Z-scores for all networks presented in the previous section. More examples can be found in Figs. 6 and 7, which show profiles for the network of flights and DBLP. Note that in the case of DBLP, the average symmetric overlap is a decreasing function of symmetric weight for the majority of samples—it is precisely this behaviour that necessitates the introduction of asymmetric measures. Results for asymmetric measures presented in both figures are qualitatively equivalent to the ones in Fig. 5. Once again, we can see a correlation between Q and v in both the actual network and the null model. A Granovetter-like relation is prominent in panel **(e)**, suggesting that the processes responsible for the distribution of weights in this network prefer to assign higher weight values to edges characterised by higher overlap values. This observation holds true for all the networks examined in our study (correlation profiles for the remaining networks are available in Supplementary Figs. S1–S5 available online).

There is another way to test Granovetter's theory—it is possible to calculate the correlation between overlaps and weights for the null model and the actual network. If Granovetter's theory is correct, then correlations in the real network should be stronger than in the null model. Figure 8 shows these correlations for all networks we studied. Considering the non-linearity of data, we decided to use the Spearman correlation and calculate it for logarithms of weights and overlaps. As can be seen, in all cases, there is a stronger positive correlation between weights and overlaps in the actual network, which supports Granovetter's theory.

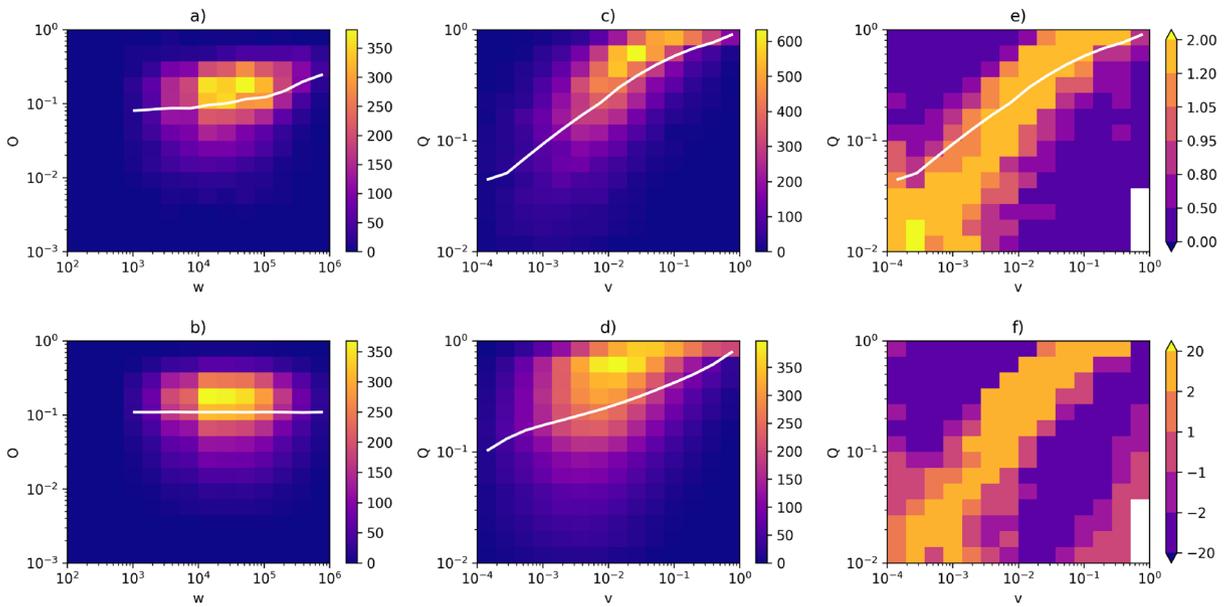


Figure 6. Correlation profiles for the network of flights. **(a)** Heatmap for the actual network—symmetric weights. **(b)** Heatmap for the null model (randomised network)—symmetric weights. **(c)** Heatmap for the actual network—symmetric weights. **(d)** Heatmap for the null model (randomised network)—symmetric weights. **(e)** Correlation profile (R). **(f)** Z-score (Z). The white lines in **(a,b)** correspond to the average O as a function of w , on **(c,d)**—the average Q as a function of v . The line in panel **(e)** is the same as in **(c)**.

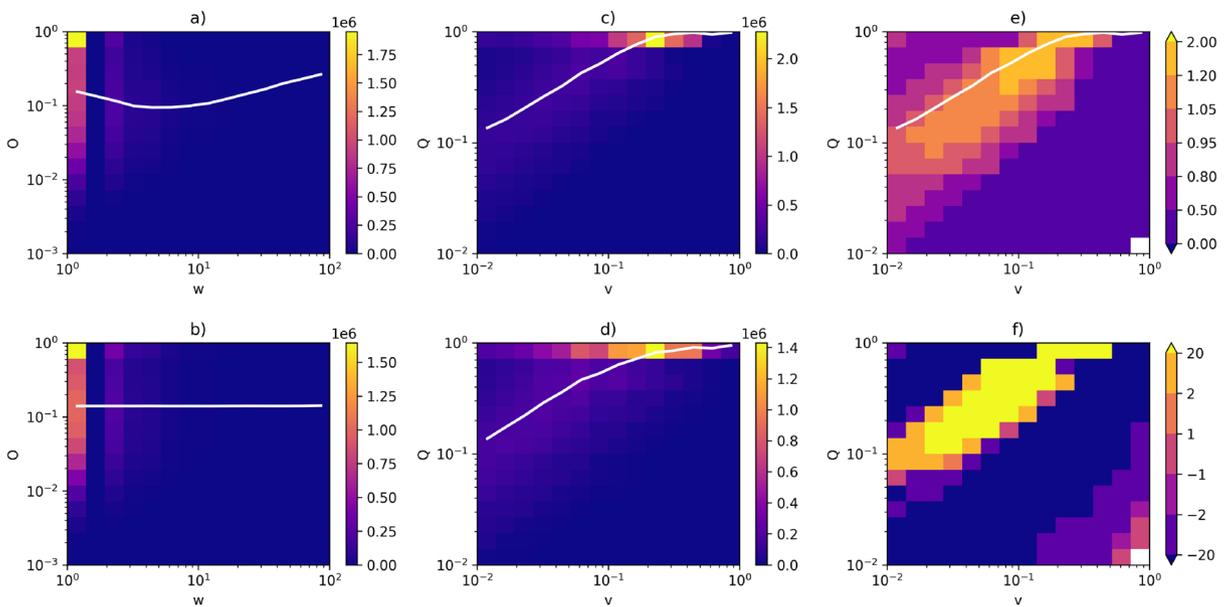


Figure 7. Correlation profiles for DBLP. **(a)** Heatmap for the actual network—symmetric weights. **(b)** Heatmap for the null model (randomised network)—symmetric weights. **(c)** Heatmap for the actual network—symmetric weights. **(d)** Heatmap for the null model (randomised network)—symmetric weights. **(e)** Correlation profile (R). **(f)** Z-score (Z). The white lines in **(a,b)** correspond to the average O as a function of w , on **(c,d)**—the average Q as a function of v . The line in panel **(e)** is the same as in **(c)**.

Summary and concluding remarks

Due to the asymmetric nature of many human interactions (or, more generally, any interactions), symmetric measures cannot be universally used to describe social networks^{14,31}. As we have shown, asymmetry is required in order to deal with such networks properly. For example, asymmetric measures can be used to confirm Granovetter’s theory in the network of scientific collaborations, which was considered a counterexample to said theory. However, asymmetric measures—depending on their definitions—are not easy to interpret and require careful and deliberate handling.

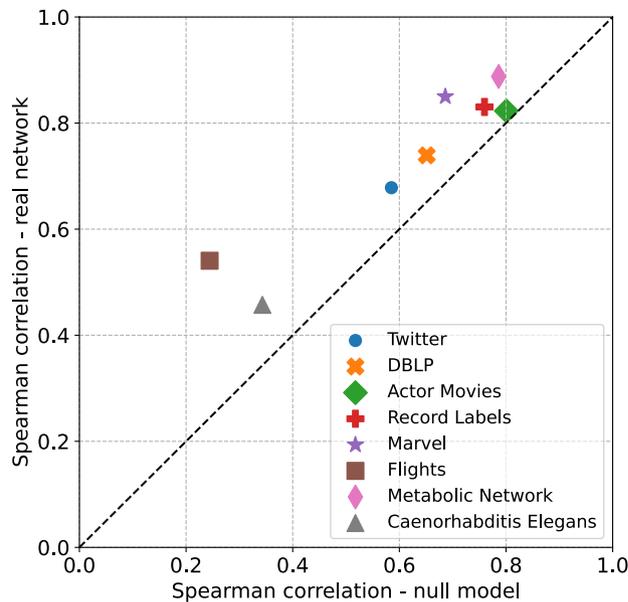


Figure 8. Spearman correlation between asymmetric overlaps and asymmetric weights in the real network as a function of the corresponding correlation in the null model.

In the case of the asymmetric overlap Q and asymmetric weight v , as defined in Eqs. (6) and (12), the problem with interpretation stems from the superfluous correlations introduced by the definitions of these measures. In fact, there are two layers of correlation that one needs to be wary of when analysing the relationship between Q and v . The first layer is purely structural, induced by the network's topology. The strength of a node s (the sum of weights over edges connecting the node to its neighbours) is correlated with the node's degree, resulting in a correlation between Q and v . The second layer of correlations, the one we are truly interested in when confirming Granovetter's theory, is tied to the sociological processes that govern the distribution of weight between edges in the network. We assume that higher weight values will be assigned to edges with higher overlap values, which is not obvious, unlike the previous correlation. The problem is that correlations from both sources overlap, and a method that would allow us to differentiate between them is needed.

In this paper, we have shown that correlation profiles can be used to achieve this goal. The idea behind them is simple but effective—by randomising weights in a graph (shuffling them), we destroy the second kind of correlations, leaving only the structural correlations intact. Then, by comparing weights in the actual graph with its randomisations, we can determine how exactly the sociological processes responsible for weight distribution in a given network assign weights to edges. Our analysis shows that in the network we studied, a clear Granovetter-like relationship is present in the correlation diagrams (see Fig. 5e for Twitter and Fig. 6e for the network of flights). That is, higher weight values are assigned, on average, to edges with higher overlap values—to the point that a monotonic relation (in the average sense on a log-log plot) is visible in the diagrams. This result truly confirms Granovetter's theory.

Moreover, not only did we study social networks and artificial social networks, but we also calculated correlation profiles for different kinds of networks—for example, the neural network of *Caenorhabditis elegans* or the metabolic network. These networks also exhibit a Granovetter-like relation between overlaps and weights, which suggests that Granovetter's theory is a sociological manifestation of more general principles governing complex networks.

On the one hand, we believe that this result is intuitive, as one can generally expect that if two nodes share a large portion of their neighbourhoods, then the strength of the connection between these nodes will likely be high. On the other hand, we hypothesise that the recently popularised theory of hidden metric spaces^{32–35}, possibly coupled with other notions (e.g. complementarity³⁶), can provide a more formal explanation of this phenomenon. According to this theory, the topology of some networks and the values of weights can be explained by the existence of metric spaces in which these networks can be embedded—the connections in the network are determined, roughly speaking, by the positions of nodes in the hidden space. Such a structured way of determining (or explaining the topology of) neighbourhoods of nodes and edge weights likely leads to a correlation between weights and overlaps. The hidden metric space models can be applied to both unipartite and bipartite networks³⁷ (in the latter case, they successfully explain some peculiar properties of these kinds of networks), which is especially interesting from the perspective of our work, considering that many networks we studied, and many real networks in general, have a bipartite representation. Moreover, some of the numerical experiments we performed indicated that a Granovetter-like relationship (albeit weak) could be present in projections of random bipartite graphs. Thus, networks with bipartite representations may be more prone to exhibit a relation between overlaps and weight similar to those presented in this paper. It should be emphasized, however, that the explanation of Granovetter's theory by hidden metric space models is still only a hypothesis and a possible and interesting direction for future studies.

Another interesting direction for future research is tied to link prediction. Recently, it was shown that link prediction methods that take advantage of link asymmetry are superior to traditional methods based on symmetric measures¹⁷. Thus, it is natural to assume that the tools presented in this paper could be helpful in link prediction and the knowledge on the two sources of correlations—structural and sociological—could be utilised to improve prediction of hidden network connections. Unfortunately, this problem is not simple. On the one hand, different link scores used in similarity-based prediction methods that are based on Eq. (6) and/or link weights (see e.g. Tab. III in¹⁷) do not distinguish between the various possible origins of shared portions of nodes' neighbourhoods (whether two nodes share a neighbour due to sheer statistics or due to their sociological relationship). On the other hand, one could modify existing scores taking into account the non-linear dependency between both types of correlations visible in the correlation profiles. For example, one of the most promising scores defined in¹⁷ is a sum of two terms: the first term, $f_1(n_{ij}, k_i, k_j)$, is based on structural properties of nodes i and j , while the second term, $f_2(v_{zi}, v_{zj})$, takes into account weights of links between these nodes and their neighbours z . One could propose an alternative score which is based on a different functional relation between these two terms. However, we believe such a relation would be case-dependent since the shape of the correlation profiles presented in this paper differs between networks. Thus, we leave this problem for future research.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Code availability

The code that supports the findings of this study is available from the corresponding author upon request.

Received: 12 February 2024; Accepted: 28 March 2024

Published online: 03 April 2024

References

- Granovetter, M. S. The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
- Granovetter, M. S. *Getting A Job: A Study of Contacts and Careers* (University of Chicago Press, 2018).
- Onnela, J.-P. et al. Structure and tie strengths in mobile communication networks. *PNAS* **104**, 7332–7336 (2007).
- Easley, D. & Kleinberg, J. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge University Press, 2010).
- Eagle, N., Macy, M. & Claxton, R. Network diversity and economic development. *Science* **328**, 1029–1031 (2010).
- Pajejic, S. & Plenz, D. The organization of strong links in complex networks. *Nat. Phys.* **8**, 429–436 (2012).
- Grabowicz, P. A., Ramasco, J. J., Moro, E., Pujol, J. M. & Eguiluz, V. M. Social features of online networks: The strength of intermediary ties in online social media. *PLoS One* **7**, e29358 (2012).
- Szell, M. & Thurner, S. Measuring social dynamics in a massive multiplayer online game. *Soc. Netw.* **32**, 313–329 (2010).
- Szell, M. & Thurner, S. Social dynamics in a large-scale online game. *Adv. Complex Syst.* **15**, 1250064 (2012).
- Šuvakov, M., Mitrović, M., Gligorićević, V. & Tadić, B. How the online social networks are used: Dialogues-based structure of myspace. *J. R. Soc. Interface* **10**, 20120819 (2013).
- Ke, Q. & Ahn, Y.-Y. Tie strength distribution in scientific collaboration networks. *Phys. Rev. E* **90**, 032804 (2014).
- Ubaldi, E., Burioni, R., Loreto, V. & Tria, F. Emergence and evolution of social networks through exploration of the adjacent possible space. *Commun. Phys.* **4**, 28 (2021).
- Pan, R. K. & Saramäki, J. The strength of strong ties in scientific collaboration networks. *Europhys. Lett.* **97**, 18007 (2012).
- Fronczak, A., Mrowinski, M. J. & Fronczak, P. Scientific success from the perspective of the strength of weak ties. *Sci. Rep.* **12**, 5074 (2022).
- Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, 2010).
- Dorogovtsev, S. & Mendes, J. *The Nature of Complex Networks* (Oxford University Press, 2022).
- Orzechowski, K. P., Mrowinski, M. J., Fronczak, A. & Fronczak, P. Asymmetry of social interactions and its role in link predictability: The case of coauthorship networks. *J. Informetr.* **17**, 101405 (2023).
- Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *PNAS* **101**, 3747–3752 (2004).
- Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
- Maslov, S., Sneppen, K. & Zaliznyak, A. Detection of topological patterns in complex networks: Correlation profile of the internet. *Phys. A* **333**, 529–540 (2004).
- Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
- Newman, M. E. J. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).
- Fronczak, A. & Fronczak, P. Networks with given two-point correlations: Hidden correlations from degree correlations. *Phys. Rev. E* **74**, 026121 (2006).
- Litvak, N. & van der Hofstad, R. Uncovering disassortativity in large scale-free networks. *Phys. Rev. E* **87**, 022801 (2013).
- Tang, J. et al. Arnetminer: Extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08*, 990–998 (Association for Computing Machinery, 2008).
- Alberich, R., Miro-Julia, J. & Rossello, F. Marvel universe looks almost like a real social network (2002). [arXiv:cond-mat/0202174](https://arxiv.org/abs/cond-mat/0202174).
- Li, C. et al. Biomodels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.* **4**, 92 (2010).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- Newman, M. E. J. & Park, J. Why social networks are different from other types of networks. *Phys. Rev. E* **68**, 036122 (2003).
- Zhou, T., Ren, J., Medo, M. & Zhang, Y.-C. Bipartite network projection and personal recommendation. *Phys. Rev. E* **76**, 046115 (2007).
- Mattie, H., Engø-Monsen, K., Ling, R. & Onnela, J.-P. Understanding tie strength in social networks using a local "bow tie" framework. *Sci. Rep.* **8**, 9349 (2018).
- Allard, A., Serrano, M. Á., García-Pérez, G. & Boguñá, M. The geometric nature of weights in real complex networks. *Nat. Commun.* **8**, 14103 (2017).
- Serrano, M. A. & Boguñá, M. *The Shortest Path to Network Geometry: A Practical Guide to Basic Models and Applications. Elements in the Structure and Dynamics of Complex Networks* (Cambridge University Press, 2022).
- Boguñá, M. et al. Network geometry. *Nat. Rev. Phys.* **3**, 114–135 (2021).

35. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. & Boguñá, M. Hyperbolic geometry of complex networks. *Phys. Rev. E* **82**, 036106 (2010).
36. Budel, G. & Kitsak, M. Complementarity in complex networks (2023). [arXiv:2003.06665](https://arxiv.org/abs/2003.06665).
37. Kitsak, M., Papadopoulos, F. & Krioukov, D. Latent geometry of bipartite networks. *Phys. Rev. E* **95**, 032309 (2017).

Acknowledgements

This research was funded by the POB Research Centre Cybersecurity and Data Science of Warsaw University of Technology within the Excellence Initiative Program-Research University (ID-UB).

Author contributions

A.F. and P.F. conceived and planned the study, M.J.M. wrote the manuscript, M.J.M. and K.P.O. performed research, wrote simulations and analysed data, all authors analysed the results and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-58357-4>.

Correspondence and requests for materials should be addressed to M.J.M.

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

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



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

© The Author(s) 2024