# Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics

**István A. Kovács, Robin Palotai, Máté S. Szalay, Peter Csermely***
Department of Medical Chemistry, Semmelweis University, Tűzoltó str. 37-47, H-1094 Budapest, Hungary
*E-mail: csermely@eok.sote.hu

**Background:** Network communities help the functional organization and evolution of complex networks. However, the development of a method, which is both fast and accurate, provides modular overlaps and partitions of a heterogeneous network, was rather difficult.
**Methodology/Principal Findings:** Here we introduce the novel concept of community landscapes and ModuLand, an integrative framework determining overlapping network modules as hills of the community landscape and including several widely used modularization methods as special cases. As various adaptations of the concept, we developed several algorithms, which provide an efficient analysis of weighted and directed networks, and (1) determine overlapping modules with a high resolution; (2) uncover a hierarchical network structure in previously unprecedented details allowing an efficient, zoom-in analysis of large networks; (3) allow the determination of key network elements and (4) help to predict network dynamics.
**Conclusions/Significance:** The concept opens a wide range of possibilities to develop new approaches and applications including network routing, classification, comparison and prediction.

## Introduction

Hundreds of module determination methods are based on roughly the same intuitive picture identifying the network communities as the dense groups of the network, in which the network elements have a much stronger influence on each other than the rest of the network. The development of a method which translates this intuitive definition of modules into practically applicable, fast and accurate, widely usable algorithm turned out to be a very challenging problem. So far a wide variety of great ideas and powerful approaches based on very different physical or algorithmic grounds were applied in order to solve this problem. At the moment there is no 'best method' available to find network modules, and even the widely used algorithms may suffer from serious problems (ESM1 Figure S1.1, ESM1 Tables S1.1 and S1.2) [1-7], although they usually provide useful and clear dissections of the networks.

## Results

### Description of the ModuLand method family

Keeping in mind the emerging needs for an integrative approach we have developed the ModuLand framework (Figure 1 and ESM1 Figure S1.2). Considering a real network as an interacting system, the quantitative simulation of the influence (or indirect impact) of a given node on the rest of the network is an interesting problem in itself. The first step of the ModuLand framework builds up these influence functions (we call them as ‘community heaps’ from now on, see Glossary) for each and every network element. A community heap can be extended over the whole network, which would produce accurate, but slow results, so practically it is beneficial to stop the influence simulation at a given threshold. As basic examples of this technique, we developed the NodeLand, LinkLand and PerturLand community heap construction methods, which we describe in detail in the Supporting Information (ESM1). In the applications shown we mostly use the LinkLand method as a simple and convenient representative of the huge family of the possible, ModuLand framework simulation techniques. As an example, Figure 1A shows three community heaps defined over the links of the ‘network science’ co-authorship network [8]. All the three starting links highlighted by the arrows belong to widely collaborating, key players of the field, resulting large community heaps.

As the second step of the ModuLand framework, the ‘community landscape’ is generated by summing up all the community heap values of a given link of the network (that is, the values of the influence functions of all other links of the network on the given link), and consequently performing this summation for each link. In order to give a visual representation, the resulting centrality-type values are plotted vertically over a 2D representation of the network resulting in a 3D visual image of the community landscape heights as shown on Figure 1B and ESM1 Figure S1.3. Now we can see ‘hills and mountains’ of the community landscape consisting of those elements, which influence each other stronger than the rest of the network. This is exactly the intuitive definition of modules given in the first sentence of this paper.

The third and last step of the ModuLand framework identifies the modules of the network by finding the ‘hills and mountains’ of the community landscape. This is seemingly easy (we can ‘see’ them on Figure 1B), but we should not forget that the position of links in their 2D network representation of Figure 1 already reflects the information on the density and strength of their interactions. There are many different ways to construct an algorithm for identifying the hills of the community landscape, each such algorithm suitable for a range of applications, but not necessary adequate for all (Section V. and Figures S1.4 and S1.5 of ESM1). One of the most straightforward choices is the ThresholdHill method, which identifies the connected components of the community landscape above a given threshold as hills. This approach results in distinct network modules (having no overlaps) like in case of the widely used Girvan and Newman method [2]. Generally it is a rather difficult problem to choose the most appropriate value for the threshold. If we raise the detection limit too high, we will find only the largest communities, and on the other hand, if we set the detection limit too low in order to be able to see the smaller modules, than most of the large communities would merge together. This is the manifestation of the well known ‘giant-component’ problem [6,9-11]. With an other hill definiton approach (like our PeakHill methods) we can overcome this hard problem, but we may encounter some new difficulties at noisy community landscapes as we describe later.

In the case of PeakHill methods we start the identification of the modules by finding the module centers, i.e. the links (or plateaus of links), which have a local maximum height on the community landscape. We worked out two implementations of the PeakHill methods in detail, the TotalHill and ProportionalHill methods (Figure 1C and ESM1 Figure S1.2). In these methods the hills (thus the module membership values of the links) are generally overlapping in a continuous way. By finding local maxima, PeakHill methods automatically yield the number of modules, and find the small and large modules simultaneously in strong contrast with the previously described ThresholdHill approach, where one often needs special criteria to determine the threshold value (see ESM1 Figure S1.4 and Table S1.2). Although for practical purposes we suggest the use the ProportionalHill method as a representative example of the PeakHill approach, the most detailed module overlap information is achievable with the computationally more expensive TotalHill method. Therefore we show results of applying the TotalHill method on Figure 1C and Figure 2, where large segments of the network belong to at least two modules.

We note that although the PeakHill approaches we described in this paper (including the ProportionalHill method suggested above) outperform the traditional ThresholdHill approach in terms of overcoming the giant-component problem and producing overlapping modules, nevertheless, they also have their own drawbacks. When applying the PeakHill approaches on a 'noisy' community landscape, each local maximum will result a new separate (and possibly highly overlapping) module. Therefore we routinely applied a simple, yet effective post-processing step for merging the groups of extremely overlapping modules (having a correlation higher than 90%) (see Section VI. of ESM1).

As a final step of our module membership assignment methods the module membership values of the links in the network are determined. Based on this result the module membership values of the nodes are simply calculated as the sum of the membership values of their links.

**General characterization of the ModuLand method**

After introducing the structure of the ModuLand framework in this Section we will characterize the approach. In principle both the calculation of the community heaps and the determination of the community landscape hills are demanding problems, requiring specific solutions depending on the precise nature of the analyzed network. However, by constructing the community landscape, the small details of the community heaps get averaged out, therefore in practical cases fast and approximate solutions of the mentioned problems are sufficient. This is the reason why rather simple community heap construction methods (like the NodeLand method) perform well on various kinds of real-world networks. On the other hand, the module membership value of any given node is obtained as the sum of the module membership value of the links of the given node, thus the small details of the hill determination step get also averaged out. The summation of the link module membership values provides an overlapping modularization of the nodes even in the absence of an overlapping modularization of the links themselves. (A similar situation is described in ref. [12].) To summarize, we divided the very challenging problem of module determination into two likewise hard subproblems, but fortunately in most cases a relatively fast, approximate treatment of these subproblems provide sufficiently fine modularizations in the end. However, depending on the precise nature of the application, it is possible, or even advised to devise a more elaborate treatment of the subproblems of community heap and community landscape hill determinations.

Several widely used, efficient network modularization methods [2,7] can be interpreted as parts of the ModuLand framework either by identifying the underlying community heap construction method or by identifying the community landscape directly (Section IV.4. of ESM1). New modularization methods can easily be generated by taking an existing ModuLand modularization protocol, and changing any of its community heap construction, landscape generation, or hill determination methods. Additionally, former methods yielding non-overlapping modules (which can be interpreted as the application of the ThresholdHill method over an appropriate landscape) can be upgraded to overlapping modularization methods using the PeakHill module determination approach (Section IV.4. of ESM1).

Optionally, a higher level hierarchical representation of the network can also be created, where the nodes of the higher level correspond to the modules of the original network, and the links of the higher level correspond to the overlaps between the respective modules (Figure 1D, ESM1 Figure S1.2 and S1.6). This hierarchical representation can be used recursively in several steps until the whole original network is represented by noninteracting elements, allowing a fast, zoom-in type analysis of large networks (Section VII. of ESM1).

Enriching the binary, yes/no module membership assignment of many previous methods, the ModuLand method-family gives a continuous scale for the association of each link and element to all modules (ESM1 Figure S1.7). To define the number of modules of a link or element the 'effective number' of modules was introduced (see Section V.6.b. of ESM1), which is a threshold-less, continuous measure based on the effective size of support of a probability distribution [13]. Additionally, the ModuLand method allowed the definition of a large set of novel, topological measures characterizing e.g. the centrality and bridgeness of network elements and links (Sections IV. and V.6. of ESM1).

**Characterization of the overlapping modules identified by the ModuLand method-family**

The ModuLand method family, even with the simplistic NodeLand community heap construction method correctly identified the observed split of the gold-standard Zachary karate club network [14] while uncovering a third, previously identified module and several club-members in modular overlaps (ESM1 Figure S1.7).

Application of the LinkLand community heap construction method to the University of South Florida word association network [15] resulted in a set of modules having a highly heterogeneous degree, module size and module overlap distribution (ESM1 Figure S1.8), which is in agreement with earlier data (see ESM1) [3,7].

The application of the ModuLand method on the benchmark graphs of Lancichinetti et al. [16] generated over a range of parameter settings showed (Figure S1.13. and Section VI.2. of ESM1) that the identified ModuLand modules corresponded consistently to the original modules, while modules can be defined in the strong sense (where ‘strong sense’ means, at least the half of the neighboring nodes are assigned to the same module as the given node, see ref. [16]).

**Variable overlaps of modules surrounding heteronym and antagonym words in a word association network**

Extending the analysis of the gold-standard Zachary karate club network, we examined the much larger University of South Florida word association network having 10,617 elements and 63,788 links [15], which was a target of a successful modularization study yielding overlapping modules [7]. This detailed analysis took 10 minutes on a computer with a 3 GHz Intel CPU. Figure 2 shows the modular environment of the antagonym word, “terrific” and that of the heteronym, “content”. The mingling colors indicate a high overlap between the modules. Importantly, the overlap of the modules with alternative meanings of the two words is much greater in the case of “terrific” than in case of “content”, which is a reasonable consequence of the fact that variations of antagonistic meanings (“terrific”) are often amongst our associations, while associations between differently pronounced meanings (“content”) are much more seldom. Overlap between the multiple meanings of the words “bright” and “focus” (ESM1 Figure S1.9) is closer to that of “terrific” than that of “content”. However, in case of these latter, multiple meaning words the similarly pronounced meanings are not divided into two major sections as in case of the antagonyms or heteronyms, which is again in agreement with our common knowledge.

**Modular hierarchy of a social network**

The modular hierarchy of the high school friendship Community-44 of the Add-Health dataset [17] was uncovered using several community heap construction methods all revealing four well-distinguishable main modules with a large amount of further sub-modules (Figure 3A and ESM1 Figures S1.10-S1.12). Girls were less likely to form multiracial friendship communities (chi-square $p < 0.05$; Figure 3B), and boys were in the overlap of significantly more friendship communities than girls (chi-square $p < 0.0001$; Figure 3C). These differences are in agreement with the sociological observations indicating a larger cohesiveness of friendship circles of girls than that of boys [18,19].

**Efficient determination of central, key elements of power-grid network**

To test whether the ModuLand framework can identify key network elements, we calculated the change of network integrity [20] during the disintegration of the USA Western Power Grid network [21]. Elements were removed in the decreasing order of their degree, betweenness centrality and ModuLand bridgeness (measuring the bridge-like role of the elements between the modules as defined in Section V.6.d. of ESM1).

Figure 4 shows that the impact of bridgeness-based element removal on network integrity was larger than that of the degree-based attacks and was well comparable to, or better than the result of betweenness centrality-based element removal. The equal-to-better performance of bridgeness-based disintegration compared to that using betweenness centrality is surprising all the more, since the global network integrity measure corresponds extremely well to the global betweenness centrality measure [20].

### Discrimination between date- and party-hubs

Discrimination of date- and party-hubs of protein interaction networks, i.e. proteins sequentially or simultaneously interacting with a large number of neighbors, is a rather difficult task [22-27]. We hypothesized, that among date-hubs and party-hubs of similar centrality, date-hubs may have a higher bridgeness (i.e. they are more overlapping between modules of the network). This assumption was substantiated by the inter-modular position of date-hubs [24,26] and by the similarly high efficiency of bridgeness-based and date-hub-based network disintegration (cf. Figure 4 with Figure 2 of [24] and [27]). The identification of the overlapping modules of a high-confidence yeast protein-protein interaction network [28] resulted in a number of modules with well-known functions (Figure 5A and ESM1 Figure S1.14). We calculated the bridgeness and centrality measures of the individual proteins, and plotted these values on Figure 5B. The separation of date- and party-hubs represented by the line of Figure 5B classified 84 party-hubs correctly of the total of 201, and 307 date-hubs of the total of 318. This result becomes even more convincing, if we consider that 10 out of 11 incorrectly identified date-hubs (91%) and 89 out of 117 incorrectly identified party-hubs (76%) have been potentially misclassified, if comparing them to the consensus of classifications [22-26]. In conclusion, by the help of the novel measures of the ModuLand-based analysis, we were able to discriminate between date- and party-hubs, thus predicting the dynamic behavior of network elements using only the topological information of their network.

## Discussion

The ModuLand method-family we introduced in this paper and in part in an earlier patent application [29] is a novel, fast and robust approach, which can be tailored for the special needs of the experimenter as well as for the conditions of the network studied. The method gives a comprehensible, hierarchical representation of large, real-world networks. Several key steps and especially their combination in the ModuLand method-family are novel, since (a) such a large variety of community heap-determination methods have not been integrated in any modularization methods; (b) community landscapes and their hills have never been used to determine network modules. Previous methods using local community detection or yielding overlapping modules (ESM1 Table S1.2) [4,7] used only one or another of the approaches presented here, and did not combine any of them to community landscapes. Hinneburg and Keim [30] used the density function landscape to determine non-overlapping clusters, but did not calculate the overlaps based on the hill detection defined in ModuLand method family. Previous network landscape methods utilized local elements of topology [23,31], while the ModuLand method assesses a wide range of structural information. Moreover, none of the previous authors used their landscapes for module determination. The recent work of Roswall and Bergstrom [5] published during the course of the current study [29] used the probability flow of random walks to construct a map of scientific communication. This method is similar to our PerturLand community heap construction method, but its application in [5] yields non-overlapping modules. The idea of determining node modules based on the link modules was also used by recent published methods, such as Ahn et al. [12].

The extensive and rich overlaps, network hierarchy, as well as the novel centrality and bridgeness measures uncovered by the ModuLand method can be used for the identification of long-range, stabilizing weak links, for the determination of the recently described creative, trend-setting elements governing network development and evolution [32], for prediction of missing links or elements, for network classification and for the design of efficient information transfer to name only a few of the many possibilities. Module overlaps might play a key role in the disconnection and synchronization of modules of complex systems, and their re-assembly during and after crisis, respectively. We invite our colleagues to design novel versions of the framework we gave, and to explore the above and other examples.

## Materials and Methods

### Networks

*Network science co-authorship network.* The giant component of the undirected, un-weighted network science co-authorship network contained 379 elements and 914 links [8]. *Karate club social network.* The weighted and undirected social network of a karate club has been reported by W. Zachary [14] containing 34 elements and 78 links. As the members of the karate club have split into two factions later, the network became a gold-standard of module determination methods [1-5,7]. *Word association network.* The giant component of Appendix A of the University of South Florida word association network (http://www.usf.edu/FreeAssociation/) [15] with removed link directions contained 10,167 elements and 63,788 weighted links, where weight refers to the association strength (see Section I.3. of ESM1). *School friendship network.* The giant component of the high school friendship Community-44 of the Add-Health database (http://www.cpc.unc.edu/projects/addhealth) [17] with removed link directions contained 1,127 elements and 5,096 weighted links, where weights represent the strengths of friendships (see Section I.4. of ESM1). *Power-grid network.* The un-weighted and undirected network of the USA Western Power Grid [21] contained 4,941 elements and 6,594 links. *Yeast protein-protein interaction network.* The giant component of the un-weighted and undirected yeast protein-protein interaction network [28] contained 2,444 elements and 6,271 links, covering approximately half of the yeast genome and the most reliable ('strongest') ~3% of the expected number of total links.

**Brief description of the representative LinkLand community heap construction method**

We give a detailed description of the different community heap construction methods in the Supporting Information (Section IV of ESM1). Here we introduce the major steps of the representative LinkLand community heap construction method utilized in each of the modularization examples shown in this paper.

The LinkLand method is a fast, but approximating method for the determination of the community heaps in weighted, undirected networks. Here the community heap belonging to the starting element or link is determined by a network walk. The starting link (and later its growing community heap) is extended by those neighboring elements and their links linking them to the existing community heap and also to each other, which will at least not decrease the 'community heap-threshold' of the existing community heap; the community heap is ready once such extension is no longer possible. The community heap-threshold of the LinkLand method is defined as the summarized weight of the links in the community heap divided by the number of nodes in the heap.

The following pseudo code shows how the LinkLand community heap construction method selects the members of the community heap in case of a given starting link of the network. The definition of the important variables used in the algorithm:

- *startLink*: the starting link of the actual community heap.
- *heapNodeList*: elements of the community heap (initially empty).
- *heapLinkList*: links of the community heap (initially empty).
- *tempList*: elements to be added to the community heap in the next round
- *actualHeapThreshold*: sum of the weight of all links in *heapLinkList* divided by the number of elements in *heapNodeList*.

```
clear tempList
add the two end-elements of startLink to tempList while tempList is not empty {
        add all elements of tempList to heapNodeList.
        for each link e connected to any elements of tempList {
           if endpoints of e are already in heapNodeList { add e to heapLinkList }
        }
        clear tempList
        recalculate actualHeapThreshold
        maxNewHeapThreshold := actualHeapThreshold

        for each element n not in heapNodeList but having non-zero links lks with an endpoint in
        heapNodeList {
           newHeapThreshold := sum of the weight of all links in heapLinkList + sum weight of link in lks
           newHeapThreshold := newHeapThreshold / (number of elements in heapNodeList + 1)
           if newHeapThreshold > maxNewHeapThreshold {
                   clear tempList
                   maxNewHeapThreshold := newHeapThreshold
           }
           if newHeapThreshold = maxNewHeapThreshold { add n to tempList }
        }
}
```

In the end of the LinkLand algorithm we find the links and elements of the community heap in the *heapLinkList* and *heapNodeList*, respectively. Identifying the community heap of one link in the LinkLand algorithm is structurally similar to a breadth-first search, therefore the runtime complexity of the algorithm is O(e(n+e)), where n is the number of nodes and e is the number of links in the network. However in practice the algorithm is very fast as a community heap of any given link rarely covers the whole network.

For downloading the ModuLand program package including the LinkLand community heap construction method see our homepage: http://www.linkgroup.hu/modules.php

**Brief description of the ProportionalHill module determination method**

We give a detailed description of the different methods determining modules based on the community landscape in the Supporting Information (Section V of ESM1). Here we show the major steps of the representative ProportionalHill module membership assignment method utilized in many modularization examples shown in this paper.

In the ProportionalHill module membership assignment method links of the network are assigned to modules of their non-lower neighboring links in the proportion of the absolute community landscape height of the respective neighboring links. The links having no higher neighboring link are assigned with full height to the respective modules defined by themselves.

At the start of the ProportionalHill module membership assignment method all links are marked as unassigned. After this, multiple rounds of link-assignments are performed: in each round, links are assigned to modules based on the assignment of previously assigned links. In each round, we descend to next slice of links, starting from the top community landscape slice, where a community landscape slice is formed by all links having the same community landscape height.

Here we describe the steps of a single round of the ProportionalHill module membership assignment method:

- The first step: each of the hill-tops/highlands of the community landscape (connected components of the actual community landscape slice without higher neighboring links) becomes a new module-core. Each link of all these connected components are assigned to their respective new modules with an assignment-strength of their full community landscape height.
- In consecutive steps, unassigned links of the community landscape slice having at least one neighboring link already assigned to the growing modules, are assigned to modules proportional to the assignment-strength of their neighbors already assigned to existing modules. In such a step, links assigned in the current step are not considered as ‘assigned neighbors’ during the respective step. The step described here is repeated until there are any unassigned links remaining in the actual community landscape slice. Once all links of the actual community

landscape slice have already been assigned to modules, the round is over and the next round begins, unless there are no more (lower) community landscape slices left, in which case the whole assignment procedure ends.

As an outcome of the ProportionalHill module membership assignment process, for each link the sum of the assignment-strength values of the given link to the different modules is equal to the community landscape height of that link.

The Linux-based computer programs of ModuLand-related methods (including the ProportionalHill method) and a Windows-based virtual Linux environment including ModuLand-related programs can be downloaded from here: www.linkgroup.hu/modules.php.

**Brief description of the TotalHill module determination method**

Here we give a brief description of the TotalHill module membership assignment method utilized in some modularization examples shown in this paper. In the TotalHill module membership assignment method the assignment of module-cores is performed as described previously for the ProportionalHill module membership assignment method, but when assigning a non-core link to modules of the neighboring links in proportion of the community landscape height of the neighboring links, the neighboring links of both non-lower and lower community landscape height are considered. This module membership assignment method is especially important, since it yields the most detailed information of the network module structure. More details and the exact algorithm can be found in the Supporting Information (Section V.2.c of ESM1).

## Glossary

*Community heap:* Based on the direct interactions in the network we calculate the effective, indirect impact of the starting element to the rest of the network. In our work we represent this new indirect interaction as a property of the original links of the network, so we calculate a new weight, the so-called community heap value for every link from a given starting element. The community heap is a connected subgraph, surrounding the starting element in which the community heap values are all larger than zero.

*Community landscape:* Integrating all community heap values, which were assigned to a given link, we get a centrality-type value called as the community landscape height of that link. The community landscape height shows that how much the given link is affected by the integrated indirect impact of all the starting elements of the network. The community landscape is defined as the sum of the community heap values. We usually represent the community landscape as a 3 dimensional image of the original network, where the horizontal plane is a 2 dimensional, ‘usual’ representation of the network, while on the vertical axis the community landscape values of network links are plotted.

## Supporting information

*Electronic Supplementary Material S1 (ESM1):* This supporting information contains a detailed description of the ModuLand method including the pseudo-codes of all algorithms used, 14 Supplementary Figures, 3 Supplementary Tables (with 18 module definitions, 129 different modularization methods, 13 module comparison methods), a Supplementary Discussion and 370 references.

## Acknowledgments

We thank Gábor Szuromi and Balázs Zalányi for help in the analysis of networks, members of the LINK-group (www.linkgroup.hu) for discussions and Tamás Vicsek for his seminar in our lab on 16th June 2005 giving us the starting encouragement to work on the ideas of this paper and for his continuous suggestions. This research was supported by research grants from the EU (FP6-518230), Hungarian Science Foundation (OTKA K-69105) and by an unrestricted grant from Unilever Hungary to the Hungarian Student Research Foundation, which helped the research of the authors.

## Author contributions

I.A.K. conceived and designed most of the ModuLand method, performed part of the network analysis and wrote part of the manuscript of the paper, M.S.S. and R.P. helped to formulate details of the method, designed the final computer programs, performed part of the network analysis and wrote part of the manuscript of the paper, P.C. gave the basic idea, suggested the network examples, helped the interpretation of the data and wrote part of the manuscript of the paper. I.A.K., M.S.S. and R.P. started their research as members of the Hungarian Research Student Association (www.kutdiak.hu/en), which provides research opportunities for talented high school students since 1996.

## Competing financial interest

The authors declare that they have no competing financial interest.

## References

1. Fortunato S (2010) Community detection in graphs. Phys Rep 486: 75-174.
2. Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci U S A 99: 7821-7826.
3. Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D (2004) Defining and identifying communities in networks. Proc Natl Acad Sci U S A 101: 2658-2663.
4. Nepusz T, Petróczi A, Négyessy L, Bazsó F (2008) Fuzzy communities and the concept of bridgeness in complex networks. Phys Rev E 77: 016107.
5. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci U S A 105: 1118-1123.
6. Fortunato S, Barthélemy M (2007) Resolution limit in community detection. Proc Natl Acad Sci U S A 104: 36-41.

7. Palla G, Derenyi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435: 814-818.
8. Newman ME (2006) Finding community structure in networks using the eigenvectors of matrices. Phys Rev E 74: 036104.
9. Berry JW, Hendrickson B, LaViolette RA, Phillips CA (2009) Tolerating the community detection resolution limit with edge weighting. arxiv.org/abs/0903.1072.
10. Fortunato S (2007) Quality functions in community detection. Proc SPIE 6601: 660108.
11. Kumpula JM, Saramäki J, Kaski K, Kertész J (2007) Limited resolution in complex network community detection with Potts model approach. Eur Phys J B 56: 41-45.
12. Ahn Y-Y, Bagrow JP, Lehmann S (2009) Communities and hierarchical organization of links in complex networks. arxiv.org/abs/0903.3178.
13. Grendar M (2006) Entropy and Effective Support Size. Entropy 8: 169-174.
14. Zachary WW (1977) An information flow model for conflict and fission in small groups. J Anthropol Res 33: 452-473.
15. Nelson DL, McEvoy CL, Walling JW, Wheeler JW (1980) The University of South Florida homograph norms. Behav Res Meth Instrum 12: 16-37.
16. Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. Phys Rev E 78: 046110.
17. Moody J (2001) Race, school integration, and friendship segregation in America. Am J Sociol 107: 679-716.
18. Baerveldt C, van Duijn MAJ, Vermeij L, van Hemert DA (2004) Ethnic boundaries and personal choice. Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils' networks. Soc Networks 26: 55-74.
19. Clark ML, Ayers M (1992) Friendship similarity during early adolescence: gender and racial patterns. J Psychol 126: 393-405.
20. Latora V, Marchiori M (2001) Efficient behavior of small-world networks. Phys Rev Lett 87: 198701.
21. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393: 440-442.
22. Bertin N, Simonis N, Dupuy D, Cusick ME, Han JD, et al. (2007) Confirmation of organized modularity in the yeast interactome. PLoS Biol 5: e153.
23. Eckmann JP, Moses E (2002) Curvature of co-links uncovers hidden thematic layers in the World Wide Web. Proc Natl Acad Sci U S A 99: 5825-5829.
24. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430: 88-93.
25. Kim PM, Lu LJ, Xia Y, Gerstein MB (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. Science 314: 1938-1941.
26. Komurov K, White M (2007) Revealing static and dynamic modular architecture of the eukaryotic protein interaction network. Mol Syst Biol 3: 110.
27. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol 3: e59.

28. Ekman D, Light S, Bjorklund AK, Elofsson A (2006) What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? Genome Biol 7: R45.
29. Kovacs I, Csermely P, Korcsmaros T, Szalay M (2006) Method for analyzing the fine structure of networks. Patent application WO 2007/093960.
30. Hinneburg A, Keim DA. An Efficient Approach to Clustering in Large Multimedia Databases with Noise. Knowledge Discovery and Data Mining; 1998. pp. 58-65.
31. Axelsen JB, Bernhardsson S, Rosvall M, Sneppen K, Trusina A (2006) Degree landscapes in scale-free networks. Phys Rev E 74: 036119.
32. Csermely P (2008) Creative elements: network-based predictions of active centres in proteins and cellular and social networks. Trends Biochem Sci 33: 569-576.
33. Gansner ER, North SC (2000) An open graph visualization system and its applications to software engineering. Softw Pract Exp 30: 1203-1233.

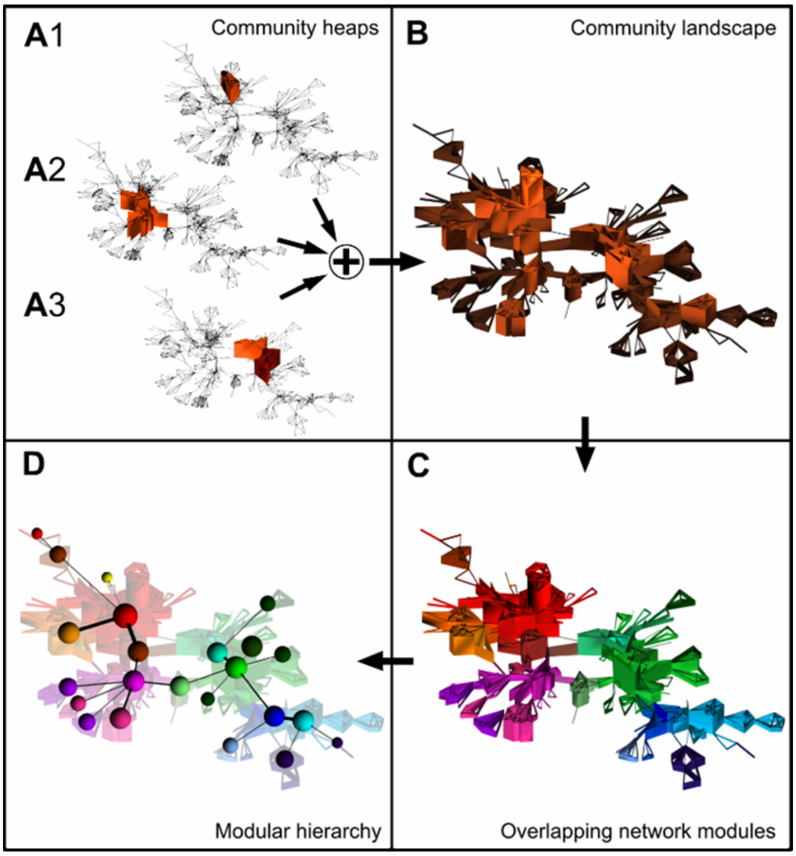


**Figure 1. Description of the ModuLand method-family.** For this illustrative example we used the network science co-authorship network [8] without link weights using the LinkLand community heap determination method with the TotalHill module membership assignment method. The network was laid out using the Kamada-Kawai algorithm and visualized with a custom Blender script. On the vertical axes community heap values (panel A), or community landscape values (panels B, C and D) of the links are shown. Community heaps of panels A1 or A2 belong to the Barabási—Vicsek or Girvan—Newman author-pairs, respectively. Panel A3 shows the merged community heap of the Arenas—Pastor-Satorras and Guimera—Amaral co-authorship links. Links and nodes of panels C and D are colored in proportion of the colors of the modules they belong. **Panel A:** community heap detection. First, the community heap of each link (or element) of the network are identified. If a link is in the 'middle' of a module, it will be a part of many community heaps (all the three widely collaborating author-pairs, whose community heap is shown by the arrows are from this category). On the contrary, links at module 'edges' will belong to few community heaps only. **Panel B:** community landscape construction. Next, the community landscape is constructed by summing up the community heap values. The hills of the community landscape correspond to the modules of the network. **Panel C:** determination of overlapping modules. Last, modular centers are identified as the links at the local maxima of the community landscapes, and memberships of links in all network modules are determined. **Panel D:** determination of network hierarchy. Optionally, a higher level hierarchical representation of the network can be created, where elements of the higher level correspond to modules of the original network, and links of the higher level correspond to overlaps between the respective modules. Sizes of higher level elements correspond to the log size of the respective lower level modules, where the module size is the sum of the membership assignment strengths of all elements to that module.

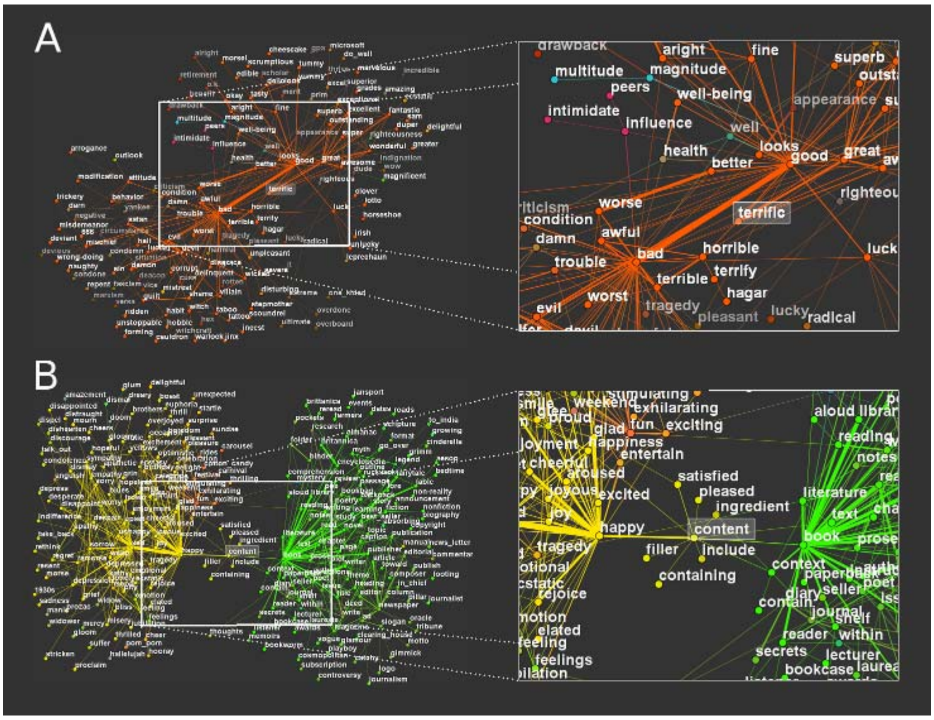


**Figure 2. Overlapping modules of a word-association network.** Modules of the University of South Florida word association network [15] were determined using the LinkLand community heap construction method and the TotalHill module membership assignment method. During the post-processing of the module assignment, we merged the modules with ProportionalHill module membership assignment-based correlation higher than 0.9 (see Section VI. of ESM1). The network was laid out using the Kamada-Kawai algorithm of Graphviz [33] and visualized using a custom program written in Python language using OpenGL graphics. Links were colored in proportion to the colors of the modules they belong. **Panel A,** Modules around the antagonym word, “terrific”. **Panel B,** Modules around the heteronym word, “content”. In addition to the selected words “terrific” and “content” similar words above a similarity threshold of 10% are also shown with a contrast corresponding to their degree of similarity. The extent of similarity between two words was calculated as the sum of the two pair-wise minima of their unity-normalized module membership vector giving the membership assignment strength of the given word to all modules of the network (for more details see Section V.6.e. of ESM1).

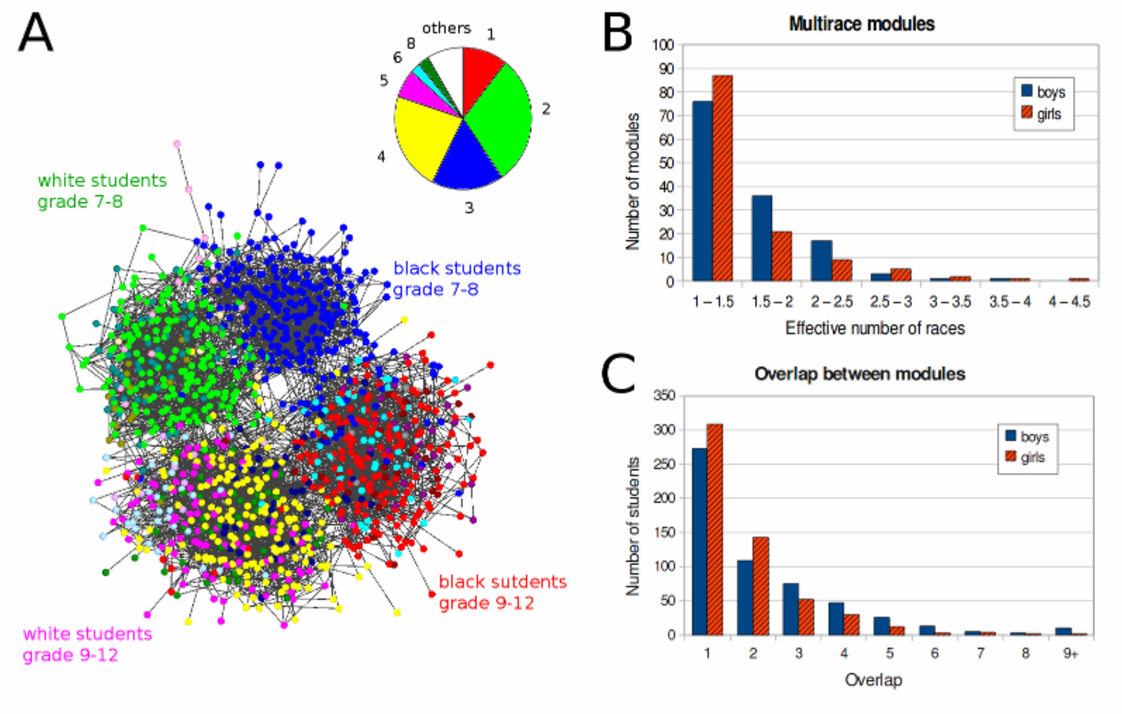


**Figure 3. Overlapping modules of a school-friendship network.** We have determined the modular structure of Community-44 of the Add Health survey [17] using the LinkLand community heap determination method together with the ProportionalHill module membership assignment method. During the post-processing of the module assignment, we merged the modules with ProportionalHill module membership assignment-based correlation higher than 0.9 (see Section VI. of ESM1). **Panel A**, Modules of Community-44. The school friendship network was laid out using the Kamada-Kawai algorithm. Elements represent the individual students, and were colored according to the color of the friendship module they assigned the most. We show the modular structure of the first hierarchical level having 18 modules. The *inset* of Panel A shows color-codes of the modules with an area proportional to the size of the respective module. **Panel B,** The number of network modules in case of boys (blue, solid bars) and girls (red-black hatched bars) with mixed racial contents at the lowest hierarchical level (level 0). The extent of mixed racial content was monitored using the 'effective number of races' (Section V.6.b. of ESM1) with a bin-size of 0.5. **Panel C,** The number of boys (blue, solid bars) and girls (red-black hatched bars) having different overlaps in friendship circles as determined in the first hierarchical level with a bin-size of 1. Overlap was measured as the 'effective number' (Section V.6.b. of ESM1) of modules of the given student.

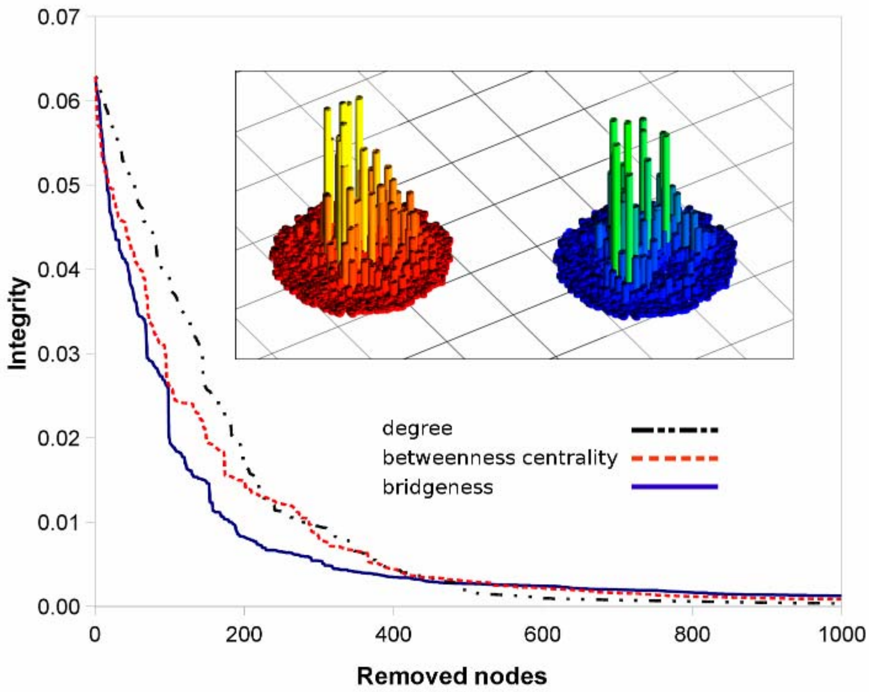


**Figure 4. Determination of key elements of the USA Western Power Grid network.** The figure shows the decreasing integrity of the USA Western Power Grid network [21] as a function of the number of elements removed. Elements were removed in the order of their decreasing degree (black alternating dashes and dots) betweenness centrality [2] (red dashed lines) or 'bridgeness' (solid blue lines), where 'bridgeness' measures the overlap of the given element between different modules as described in detail in Section V.6.d. of ESM1. Network integrity has been calculated after Latora and Marchiori [20]. Bridgeness was calculated from the modular structure of the lowest hierarchical level as determined by the LinkLand community heap construction method and the TotalHill module membership assignment method. During the post-processing of the module assignment, we merged the modules with ProportionalHill module membership assignment-based correlation higher than 0.9 (see Section VI. of ESM1). On the vertical axis of the *insets* the betweenness centrality (left, color-coded from red to yellow) and bridgeness (right, color-coded from blue to green) of the elements of the USA Western Power Grid network are shown. Networks on the *insets* were laid out using the Kamada-Kawai algorithm and visualized with a custom Blender script.

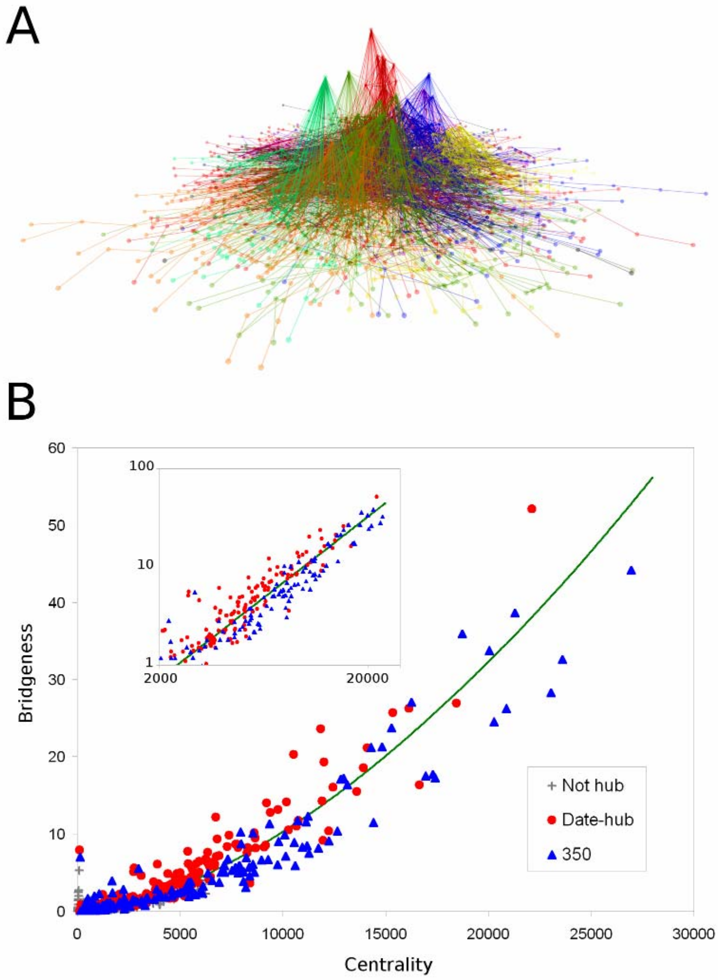


**Figure 5. Prediction of the dynamical behavior of network elements: segregation of date- and party-hubs based on their modular overlaps.** Overlapping modules of the yeast protein-protein interaction network of Ekman et al. [28] were identified using the LinkLand community heap determination method with the TotalHill module membership assignment method using the modular structure of the lowest level of hierarchy. During the post-processing of the module assignment, we merged the modules with ProportionalHill module membership assignment-based correlation higher than 0.9 (see Section VI. of ESM1). **Panel A**, 3D view of the yeast protein-protein interaction network. The underlying 2D network layout was set by the Kamada-Kawai algorithm. The vertical positions reflect the community landscape values of the elements on a linear scale. Elements were colored as the module of their maximum membership. **Panel B,** Centrality and bridgeness of yeast date- and party-hubs. Hubs having more than 8 neighbors and non-hubs with less neighbors were positioned on the scattergram according to their ModuLand centrality (x-axis, the height of the community landscape) and ModuLand bridgeness (y-axis) as defined in Section V.6.d. of ESM1. Date- and party-hubs are marked with red circles and blue triangles, respectively, while non-hub proteins are represented by gray crosses. The *inset* shows a double logarithmic plot of hubs with large centrality.