1 Application of Graph-Entropy for Knowledge Discovery and Data Mining in Bibliometric Data

André Calero Valdez, Matthias Dehmer, Bernhard Ofner, and Andreas Holzinger

1.1 Introduction

Entropy, originating from statistical physics is a fascinating and challenging concept with many diverse definitions and various applications.

Considering all the diverse meanings, entropy can be used as a measure for disorder in the range between total order (structured) and total disorder (unstructured) [1,2], as long as by "order" we understand that objects are segregated by their properties or parameter values. States of lower entropy occur when objects become organized, and ideally when everything is in complete order the Entropy value is zero. These observations generated a colloquial meaning of entropy [3].

Following the concept of the mathematical theory of communication by Shannon & Weaver (1949) [4], entropy can be used as a measure for the *uncertainty in a data set*. The application of entropy became popular as a measure for system complexity with the paper by Steven Pincus (1991) [5], who described Approximate Entropy as a statistic quantifying regularity within a wide variety of relatively short (greater than 100 points) and noisy time series data. The development of this approach was initially motivated by data length constraints, which is commonly encountered in typical biomedical signals including: heart rate, electroencephalography (EEG), etc. but also in endocrine hormone secretion data sets [6].

Hamilton et al. [7] were the first to apply the concept of entropy to bibliometrics to measure interdisciplinarity from diversity. While Hamilton et al. work on citation data, a similar approach has been applied by Holzinger et al. [8,9] using enriched meta-data for a large research cluster.

1.1.1
Challenges in bibliometric data sets, or why should we consider entropy measures?

The challenges in bibliometric data stem from various sources. First data integrity and and data completeness can never be assumed. Thus, bibliometrics faces the following problems:

- Heterogeneous data sources: need for data integration and data fusion
- Complexity of the data: network-dimensionality
- Large data sets: manual handling of the data nearly impossible
- Noisy, uncertain, missing, dirty data: careful data pre-processing necessary.

Beyond these data-integrity problems, problems of interpretation and application are important. Meyer [10] lists six stylized facts that represent recurring patterns in bibliometric data.

- Lotka's Law [11] (Frequency of publication per author in a field)
- Matthew effect: Famous researchers receive a lot more citations than less prominent researchers [12]
- Exponential growth of the number of scientists and journals [13]
- Invisible schools of specialties for every 100 scientists [13]
- Short-citation half life
- Bradfort's law of scattering of information

These stylized facts lead to the consideration of analyzing publication data using graph-based entropy analysis. Bibliometric data is similar to social network data (e.g. small world phenomenon) and obeys above-mentioned laws. In these type of network data graph-entropy may reveal potentials unavailable to standard social-network-analysis methodology.

Entropy measures have successfully been tested for analyzing short, sparse and noisy time series data. However, they have not yet been applied to weakly structured data in combination with techniques from computational topology. Consequently, the inclusion of entropy measures for discovery of knowledge in bibliometric data promises to be a big future research issue and there are a lot of promising research routes.

Particularly, for data mining and knowledge discovery from noisy, uncertain data, graph-entropy based methods may bring some benefits. However, in the application of entropy for such purposes are a lot of unsolved problems. In this chapter we will focus on the application of topological entropy and open research issues involved.

Generally, Graph theory provides powerful tools to map data structures and to find novel connections between single data objects [14, 15]. The inferred graphs can be further analyzed by using graph-theoretical and statistical techniques [16].

A mapping of aforementioned hidden schools as a conceptual graph and the subsequent visual and graph-theoretical analysis may bring novel insights on hidden patterns in the data, which exactly is the goal of knowledge discovery [17]. Another benefit of the graph-based data structure is in the applicability of methods from network topology and network analysis and data mining, e.g. small-world phenomenon [18, 19], and cluster analysis [20, 21].

1.1.2
Structure of this chapter

This chapter is organized as follows: We have already seen a short introduction into the problems of bibliometrics and how entropy could be used to tackle these problems. Next we investigate the state of the art in graph-theoretical approaches and how they are connected to text-mining (see section 1.2.1). This prepares us to understand how graph entropy could be used in data-mining processes (see section 1.2.2). Next we show how different graphs can be constructed from bibliometric data and what research problems can be addressed by each of those (see section 1.2.3). We then focus on co-authorship graphs in order to identify collaboration styles using graph entropy (see section 1.3). For this purpose we selected a subgroup of the DBLP database and prepared it for our analysis (see section 1.4). The results (see section 1.5) show how two entropy measures describe our dataset. From these results we conclude our discussion of the results and consider different extensions on how to improve our approach (see section 1.6).

1.2
State of the Art

Many problems in the real-world can be described as relational structures. Graph-Theory [22] provides powerful tools to map such data structures and to find novel connections between single data objects [14, 15]. The inferred graphs can be further analyzed by using graph-theoretical and statistical techniques [16]. A mapping of already existing and in medical practice approved *knowledge spaces* as a conceptual graph and the subsequent visual and graph-theoretical analysis may bring novel insights on hidden patterns in the data, which exactly is the goal of knowledge discovery [17]. Another benefit of a graph-based data structure is in the applicability of methods from network topology and network analysis and data mining, e.g. small-world phenomenon [18,19], and cluster analysis [20,21].

The first question is "How to get a graph?", or simpler "How to get point sets?", because point cloud data sets (PCD) are used as primitives for such approaches. Apart from "naturally available" point clouds (e.g. from laser scanners, or resulting from protein structures or protein interaction networks [23], or also text can be mapped into a set of points (vectors) in $\mathbb{R}^n$.), the answer to this question is not trivial; for some solutions see [24].

1.2.1 Graphs and Text Mining

Graph-theoretical approaches for Text Mining emerged from the combination of the fields of data mining and topology, especially graph theory [25]. Graphs are intuitively more informative as example words/phrase representations [26]. Moreover graphs are the best studied data structure in computer science and mathematics and they also have a strong relation with logical languages [25]. Its structure of data is suitable for various fields like biology, chemistry, material science and communication networking [25]. Furthermore, graphs are often used for representing text information in natural language processing [26]. Dependency graphs have been proposed as a representation of syntactic relations between lexical constituents of a sentence. This structure is argued to more closely capture the underlying semantic relationships, such as subject or object of a verb, among those constituents [27].

The beginning of graph-theoretical approaches in the field of data mining was in the middle of the 1990's [25] and there are some pioneering studies such as [28–30]. According to [25] there are five theoretical bases of graph-based data mining approaches such as (1) subgraph categories, (2) subgraph isomorphism, (3) graph invariants, (4) mining measures and (5) solution methods. Furthermore, there are five groups of different graph-theoretical approaches for data mining such as (1) greedy search based approach, (2) inductive logic programming based approach, (3) inductive database based approach, (4) mathematical graph theory based approach and (5) kernel function based approach [25].

There remain many unsolved questions about the graph characteristics and the isomorphism complexity [25]. Moreover the main disadvantage of graph-theoretical text mining is the computational complexity of the graph representation. The goal of future research in the field of graph-theoretical approaches for text mining is to develop efficient graph mining algorithms which implement effective search strategies and data structures [26].

Graph-based approaches in text mining have many applications from biology and chemistry to internet applications [31]. According to Morales et al [32] graph-based text mining approach combined with an ontology (e.g. the Unified Medical Language System - UMLS) can lead to better automatic summarization results. In [33] a graph-based data mining approach was used to systematically identify frequent co-expression gene clusters. A graph-based approach was used to disambiguate word sense in biomedical documents in Agirre et al. [34]. Liu [35] proposed a supervised learning method for extraction of biomedical events and relations, based directly on subgraph isomorphism of syntactic dependency graphs. The method extended earlier work [36] that required sentence subgraphs to exactly match a training example, and introduced a strategy to enable approximate subgraph matching. These method have resulted in high-precision extraction of biomedical events from the literature.

While graph-based approaches have the *disadvantage* of being computationally expensive, they have the following *advantages*:

- It offers a far more expressive document encoding than other methods [26].
- Data which is graph structured widely occurs in different fields such as bibliometrics, biology, chemistry, material science and communication networking [25].

A good example for graph learning has been presented by Liu et al. (2009) [37]: they proposed a graph learning framework for image annotation, where at first the image-based graph learning is performed to obtain candidate annotations for each image and then word-based graph learning is developed to refine the relationships between images and words to get final annotations for each image. To enrich the representation of the word-based graph, they designed two types of word correlations based on web search results besides the word co-occurrence in the training set. Generally, image annotation methods aim to learn the semantics of un-tagged images from already annotated images to ensure an efficient image retrieval.

1.2.2 Graph-Entropy for Data Mining and Knowledge Discovery

Rashevsky [38], Trucco [39], and Mowshowitz [40], were amongst the first researchers to define and investigate the entropy of graphs.
Graph Entropy was described by [41] to measure structural information content of graphs, and a different definition, more focused on problems in information and coding theory, was introduced by Körner in [42]. Graph entropy is often used for the characterization of the structure of graph-based systems,

e.g. in mathematical biochemistry, but also for any complex network [43]. In these applications the entropy of a graph is interpreted as its structural information content and serves as a complexity measure, and such a measure is associated with an equivalence relation defined on a finite graph; by application of Shannon's Eq. 2.4 in [44] with the probability distribution we get a numerical value that serves as an index of the structural feature captured by the equivalence relation [44].

The open source graph visualization tool Gephi allows for several different graph analyses of network graphs. Traditionally these are used with social network graphs (i.e. co-authorship graphs). Interpretation of graph statistics must be reevaluated for mixed node graphs. Graph statistics that are of interest in regard to publication networks are:

- Network entropies have been developed to determine the structural information content of a graph [45], [44]. We have to mention that the term network entropy cannot be uniquely defined. A reason for this is that by using Shannon's entropy [46], [41], [47] the probability distribution cannot be assigned to a graph uniquely. In the scientific literature, two major classes have been reported [45], [48], [49]:
  1. Information-theoretic measures for graphs which are based on a graph invariant X (e.g., vertex degrees, distances etc.) and an equivalence criterion [41]. By starting from an arbitrary graph invariant X of a given graph and an equivalence criterion, we derive a partitioning. Thus, one can further derive a probability distribution. An example thereof is to partition the vertex degrees (abbreviated as $\delta(v)$) of a graph into equivalence classes, i.e., those classes only contain vertices with degree $i = 1, 2, ..., \max \delta(v)$, see e.g. [50].
  2. Instead of determining partitions of elements based on a given invariant, Dehmer [48] developed an approach which is based on using so called information functionals. An information functional f is a mapping which maps sets of vertices to the positive reals. The main difference to partition-based measures (see previous item) is that we assign probability values to every individual vertex of a graph (and not to a partition), i.e.,

$$p^f(v_i) := \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \tag{1.1}$$

     As the probability values depend on the functional f, we infer a family of graph entropy measures

$$I_f(G) := -\sum_{i=1}^{|V|} p^f(v_i) \log p^f(v_i) \tag{1.2}$$

$|V|$ is the size of the vertex set of G. Those measures have been extensively discussed in [50].

Evidently, the both graph measures can be interpreted as graph complexity measures [50]. The latter outperform partition-based entropy measures because they integrate features from every vertex instead of subgraphs. This is important because when we look at bibliometric data (e.g. co-authorship graphs) often differ to small degrees. Measuring these with partition-based entropy could lead to highly similar data for dissimilar graph data.

1.2.3 Graphs from Bibliometric Data

Graph entropy in bibliometric data can be applied to various forms of data. Depending on how the meta-data is interpreted different types of graphs can be constructed [51]. The question that we can apply differ depending on the type of graph. One must ask: What does graph entropy mean when bibliometric graph is analyzed. For this purpose we first list various types of bibliometric graph representations.

- **Collaboration-based graphs / Co-Authorship Graphs:** In a co-authorship graph vertices represent unique authors that have published articles. Edges are inserted by connecting vertices that have published articles as *co-authors*. Edge weights can be mapped to the frequency of collaboration. A key benefit of this type of analysis is that it can be applied using meta-data alone that is publicly available. Authorship graphs are undirected graphs. Typical analyses are conducted to understand patterns of collaboration, interdisciplinarity, and the evolution of scientific subjects [52].
  - *Author level*: When edges represent individual authors, we speak about author level co-authorship graphs.
  - *Institutional level:* When edges represent institutions, we speak about institutional level co-authorship graphs.
- **Citation-based graphs:** Mapping citations from articles requires more data than often available [53]. As vertices we use articles that are joined by citation edges. Obviously, these graphs can only be constructed when citation data is available. A citation-based graph is a directed graph. No weights are assigned to edges. In these graphs analyses can be conducted in various forms. Typical analyses are co-citation analysis (i.e. what documents get cited together [54]), centrality analyses (e.g. what documents form the core of knowledge), and bibliographic coupling [51] (i.e. what documents cite similar documents). These measures are also

often used to identify scientific subject or the degree of interdisciplinarity of a journal.

  - *Article level*: When nodes represent individual articles, we speak about article level citation graphs.
  - *Journal level:* When nodes represent journals, we speak about journal level citation graphs.
  - *Subject level:* When nodes represent scientific subjects, we speak about subject level citation graphs.

- **Content/Topic-based graphs:** When full text or abstract data is available content of articles may also be used in a graph-based representation. Using different text-mining approaches topics may be identified. These can be used to map various information. Often topic-based graphs are multimodal representing relationships between different entities. These graphs are often used for recommendation purposes or to identify trends.
  - *Author-Topic mapping*: When nodes represent authors and topics, analyses can be performed to understand how authors contribute to different topics.
  - *Journal-Topic mapping:* When nodes represent journals and topics, we can analyze how topics are formed and which journals are the main contributors to a topic how they.
  - *Article-Topic mapping:* When nodes represent articles and topics, we can analyze which articles (and thus which authors) have formed a topic and how it develops over time.
- **Other/combined graphs** Using the aforementioned graphs we can factor in various forms of metadata (e.g. time-series data, citation data, etc.) to combine different approaches. For example we can use the publication date and citation data to identify how certain groups of authors have formed topics and where central ideas come from.

1.3 Identifying Collaboration-Styles Using Graph Entropy from Bibliometric Data

Bibliometrics or Scientometrics is the discipline of trying to discover knowledge from scientific publication in order to understand science, technology and innovation. Various analyses have been conducted in scientometrics using co-authorship networks as reviewed by Kumar [52]. Collaboration styles have been investigated by Hou et al. [55] identifying patterns of social proximity for the field of scientometrics itself. Topics were identified using co-occurrence analysis and collaborative fields were identified.

Applying graph entropy to publication data could be used to determine how scientific collaboration differs in various sub-fields. By analyzing co-authorship graphs in sub-communities, we could be able to identify structural differences in and between groups.

1.4 Method and Materials

In our example we want to address the most simple form of bibliometric graph data. We use this type of data to test how graph entropy works in our scenario and combine it with other methods (i.e. community detection). The aim of this approach is to identify how different communities (i.e. group of authors that co-author articles) differ in their topology.

For this purpose we evaluate the DBLP database of computer science. The XML database contains meta-data on publications in the field of computer science and covers over 3 mio. articles (as of Sep. 2015). In order to limit computation times, we focus on data only from the largest journals that deal with graph theory (see Tab. 1.1).

Because we are interested in the structure of a collaboration graph, we focus on measures that account for symmetry in the graph. Topological information content is used to measure local symmetry within communities and parametric graph entropy is used to measure the overall symmetry of a sub-community. By reviewing the influence of both we see how symmetry plays out from a detail- and meta-perspective.

Journal Name	**Articles**
Graphs and Combinatorics	710
SIAM J. Discrete Math.	644
Ars Comb.	641
IEEE Transactions on Information Theory	489
Discrete Applied Mathematics	480
Electronic Notes in Discrete Mathematics	432
J. Comb. Theory, Ser. B	412
SIAM J. Comput.	402
Combinatorics, Probability & Computing	327
IEEE Trans. Knowl. Data Eng.	277

Tab. 1.1 Largest 10 Journals from DBLP and their article count

For all years about 4,811 publications were present. By extracting author names and constructing a co-authorship graph, we get a network of 6,081 vertices (i.e. authors) and 8,760 edges (i.e. collaborations). No correction for multiple author-names were performed. Duplicate entries with different spellings are considered as two distinct entries. We could remove duplicates by applying similarity measures (e.g. Levensthein distance, etc.), but for our approach this is not necessary. For community-detection we ran the Louvain algorithm supplied by the igraph R package. The algorithm returned 1347 distinct communities and a modularity of 0.93.

After sorting communities we measure topological information content to determine the characteristics of collaboration in these sub-communities. We evaluated the following graph entropies:

1. A partition-based graph entropy measure called *topological information content* based on vertex orbits due to [41].

2. Parametric graph entropies based on a special information functional f due to Dehmer [48]. The information functional we used is

$$f(v_i) := \sum_{k=1}^{\rho(G)} c_k |S_k(v_i, G)|, \text{ with } c_k > 0 \tag{1.3}$$

summing the product of both the size of the k-sphere (i.e. the amount of nodes in G with a distance of k from v_i given as $|S_k(v_i, G)|$) and arbitrary positive correction coefficients c_k for all possible k from 1 to the diameter of the graph G. The resulting graph entropies have been defined by

$$I_f := -\sum_{i=1}^{|V|} p^f(v_i) \log p^f(v_i) \tag{1.4}$$

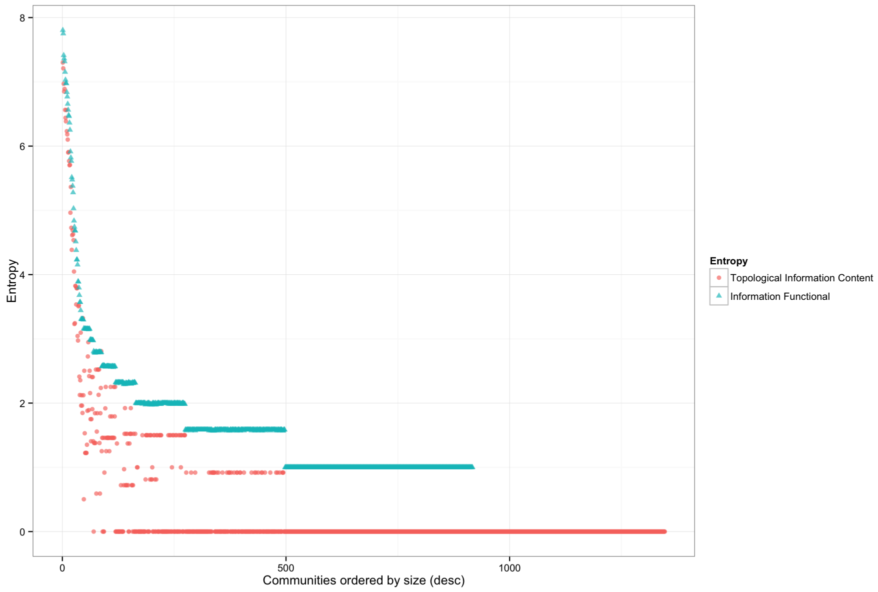
1.5 Results

The largest ten communities range from a size of 225 to 115. In order to identify communities we measure Eigenvector-Centrality for the identified sub-communities and identify the top-3 most central authors (see 1.2). We then determine our two entropy measures for the given sub-communities to characterize the collaboration properties within these communities.

We can note that the used graph entropies evaluate the complexity of our communities differently (see Fig. 1.3 and Fig. 1.4). Both seem to plot logarithmic curves but show different dispersion from an ideal curve. The distribution plot also shows typical properties of bibliometric data. Both entropies follow a power law distribution (see Fig. 1.3). The topological information content seems to scatter more strongly than parametric graph entropy. On the other hand the information functional based graph entropy seems to follow the steps of the community-size more precisely.

ID	Size	Most-Central 3 Authors	I_{mowsh}	I_{dehm}
1	225	Noga Alon; Alan M. Frieze; Vojtech Rödl	7.3	7.804
2	217	Douglas B. West; Ronald J. Gould; Alexandr V. Kostochka	7.21	7.751
3	172	Daniel Král; Ken-ichi Kawarabayashi; Bernard Lidický	6.97	7.414
4	166	Muriel Médard; Tracey Ho; Michelle Effros	6.849	7.363
5	161	Hajo Broersma; Zsolt Tuza; Andreas Brandstädt	6.889	7.317
6	141	Xueliang Li; Cun-Quan Zhang; Xiaoyan Zhang	6.563	7.153
7	132	Hong-Jian Lai; Guizhen Liu; Hao Li	6.44	7.031
8	127	Syed Ali Jafar; Abbas El Gamal; Massimo Franceschetti	6.385	6.98
9	127	Michael A. Henning; Ping Zhang; Odile Favaron	6.561	6.975
10	115	Jayme Luiz Szwarcfiter; Celina M. Herrera de Figueiredo; Dominique de Werra	6.232	6.838

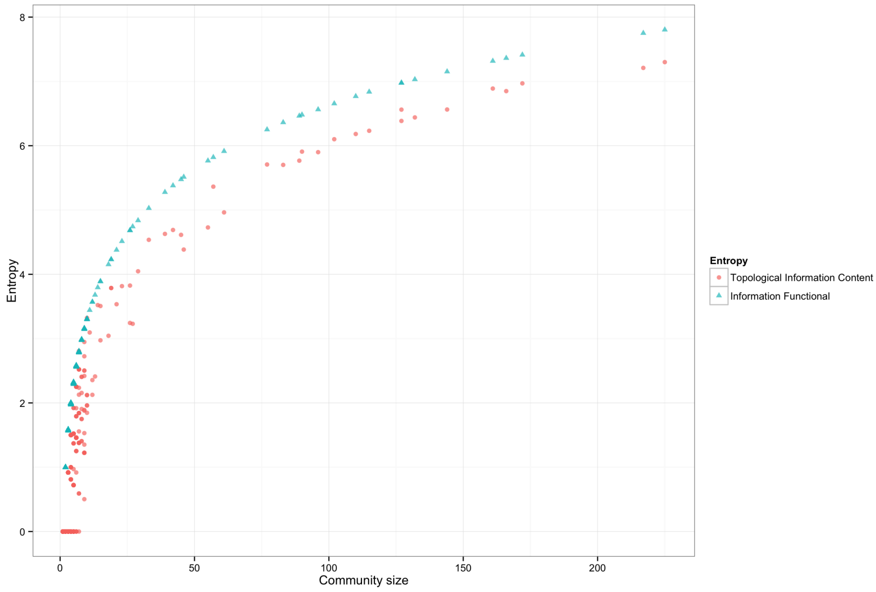
Tab. 1.2 Largest 10 Identified Communities



Tab. 1.3 Toplogical information content and parametric graph entropy distributions.

Now we will explore this problem with an example, namely by considering the measures $I_{\text{mowsh}} < I_{\text{dehm}}$ for the largest sub-communities (i.e. $ID = 1$). In this context, the inequality $I_{\text{mowsh}} < I_{\text{dehm}}$ can be understood by the fact those entropies have been defined on different concepts.

As mentioned, I_{mowsh} is based upon the automorphism group of a graph and, therefore, can be interpreted as a measure of symmetry. This measure

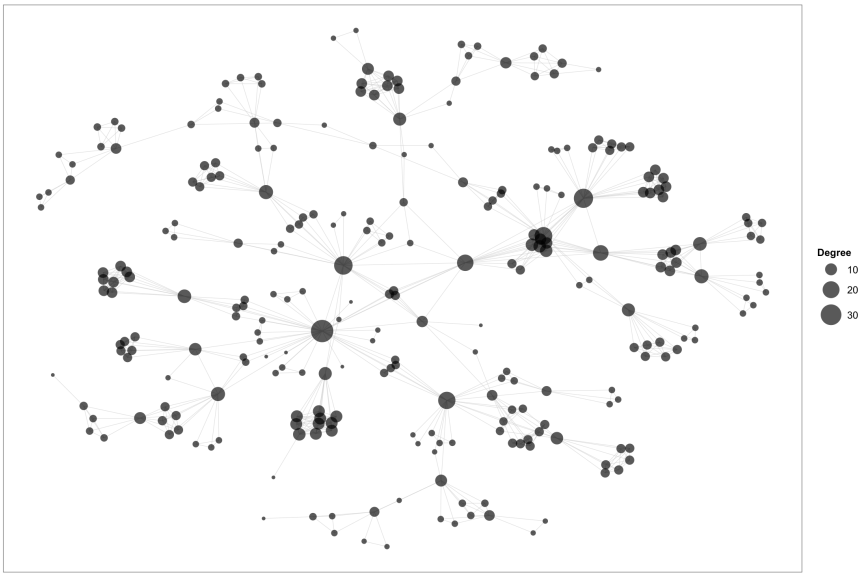


Tab. 1.4 Both entropies plotted over community size.

becomes small when all vertices are located in only one orbit. By contrast, the measure is maximal ($= \log_2(|V|)$) if the input graph equals the so-called identity graph; that means all vertex orbits are singleton sets. In our case, we obtain $I_{\text{mowsh}} = 7.3 < \log_2(225) = 7.814$ and conclude that according to the definition of I_{mowsh}, the community is rather symmetrical.

Instead, the entropy I_{dehm} characterizes the diversity of the vertices in terms of their neighborhood, see [45]. The higher the value of I_{dehm}, the less topologically different vertices are in the graph and, finally, the higher is the inner symmetry of our sub-community. Again, maximum entropy for our network equals $\log_2(225) = 7.814$. Based on the fact that for the complete graph K, $I_{\text{dehm}}(K_n) = \log(n)$ holds, we conclude from the result $I_{\text{dehm}} = 7.804$ that the community network is highly symmetrical and connected and could theoretically be obtained by deleting edges from K_{225} (see also Fig. 1.5). A similar conclusion can be derived from looking at $I_{\text{mowsh}} = 7.3$.

In comparison the values of community $ID = 30$ differs regarding these values. Its topological information content is $I_{\text{mowsh}} = 2.355$, while its parametric graph entropy is $I_{\text{dehm}} = 3.571$. The theoretical maximum for this graph is $log_2(12) = 3.58$ — very near to the parametric graph entropy. When looking at the resulting network plot (see Fig. 1.6) we can see that the graph is symmetrical on a higher level. We have three sub-communities, all held to-



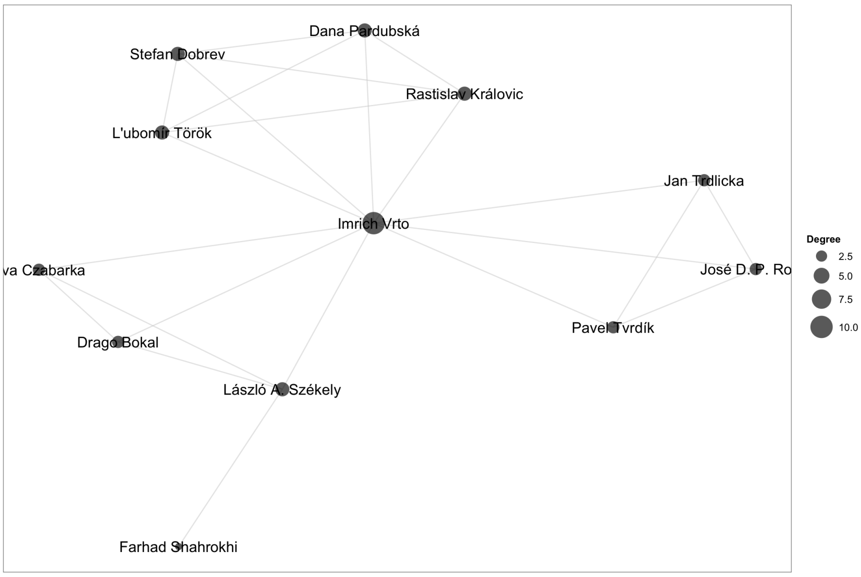
Tab. 1.5 Plot of the largest community in the co-authorship graph.

gether by the central author "Imrich Vrto". The graph is thus less symmetrical on a higher order, but the inner symmetry is still high.

1.6 Discussion and Future Outlook

Different entropy measure deliver different results because they are based on different graph properties. When using the aforementioned entropy measures in a co-authorship graph measures of symmetry I_{dehm} (based on vertex neighborhood diversity) or I_{mowsh} (based on the graph automorphism) deliver different measures of entropy. Interpreted we could say, that authors can be similar in regard to their neighborhoods (i.e. authors show similar publication patterns) while the whole graph shows low measures of automorphism-based symmetry to itself. This could mean authors can not be exchanged for one another without changing basic properties of the graph. On the other hand when I_{mowsh} is significantly lower than I_{dehm} we could argue that symmetry differs on different levels of the graph. Interpreting these differences could be more interesting than looking at the individual symmetry measures themselves.

Since publications and thus collaboration are time related, one could extend this approach to Markov-networks. Applying various graph-entropy mea-



Tab. 1.6 Plot of the 30th community in the co-authorship graph with author names.

sures in this context could reveal changes in collaboration and indicate a shift in topics for authors or subgroups of authors.

1.6.1
Open problems

From our work we must say, that deriving co-authorship communities based on Louvain-clustering naturally leads to specific structures in community building. The created communities are probabilistic estimates of real communities. The investigated communities tend to show high similarity for the parametric graph entropy. This is expected, as they are constructed by removing edges from the full graph that is separated into sub-graphs that should be coherent clusters. Our analysis shows that we can derive properties from co-authorship graphs that represent collaboration behavior, but our method is biased. It is likely to fail with small collaboration groups, as their entropy can not take up that many different values. One approach to tackle this problem could be to used bimodal graphs that include publication nodes. This however, leads to drastically larger graphs, which in turn require more processing power.

For further investigations one could use empirical data or integrate text-mining approaches to identify more accurate clusters. Using non-exclusive clustering methods could also improve on our results. Additional measures of entropy should also be used to evaluate found communities.

1.6.2 A Polite Warning

Bibliometric analyses tend be used in evaluations of scientific success quite often. Sadly, they are often used with only introductory knowledge in bibliometric evaluation. The purpose of this chapter is not to propose a method for evaluating research performance, but to provide new methods for the analysis of collaboration. Major deficits in this approach for performance measurement stem from typical bibliometric limitations (e.g. database coverage, author identification, etc.). Using these methods for performance evaluations without considering these limitations reveals a lack of understanding of bibliometrics and should therefore be left to bibliometric experts.

Bibliography

1 Holzinger, A. In *DATA 2012*, INSTICC, Rome, Italy, **2012**, pp. 9–20

2 Holzinger, A., Stocker, C., et al. In *Communications in Computer and Information Science CCIS 455*, (eds. M. S. Obaidat, J. Filipe), Springer, Berlin Heidelberg, **2014**, pp. 3–18

3 Downarowicz, T. *Entropy in dynamical systems*, vol. 18, Cambridge University Press, Cambridge, **2011**

4 Shannon, C. E., Weaver, W. *The Mathematical Theory of Communication*, University of Illinois Press, Urbana (IL), **1949**

5 Pincus, S. M. *Proceedings of the National Academy of Sciences* 88 (**1991**), pp. 2297

6 Pincus, S. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 5 (**1995**), pp. 110

7 Hamilton, K. S., Narin, F., & Olivastro, D. (2005). Using bibliometrics to measure multidisciplinarity. Westmont, NJ

8 Holzinger, A., Ofner, B., Stocker, C., Calero Valdez, A., Schaar, A. K., Ziefle, M., & Dehmer, M. (2013). On graph entropy measures for knowledge discovery from publication network data. In Availability, reliability, and security in information systems and HCI (pp. 354-362). Springer Berlin Heidelberg.

9 Calero Valdez, A., Schaar, A. K., Ziefle, M., Holzinger, A., Jeschke, S., & Brecher, C. (2014). Using mixed node publication network graphs for analyzing success in interdisciplinary teams. In Automation, Communication and Cybernetics in Science and Engineering 2013/2014 (pp. 737-749). Springer International Publishing.

10 Meyer, M. (2011) In *Journal of Artificial Societies and Social Simulation*, 14(4), **2011**

11 Lotka, A. J. (1926). The frequency distribution of scientific productivity. Journal of Washington Academy Sciences, Vol. 16, No. 12, pp. 317-324.

12 Merton, R. K. (1968). The Matthew effect in science. Science, Vol. 159, No. 3810, pp. 56-63.

13 De Solla Price, D. J. (1963). Little science, big science... and beyond. New York: Columbia University Press.

14 Strogatz, S. *Nature* 410 (**2001**), pp. 268

15 Dorogovtsev, S., Mendes, J. *Evolution of networks: From biological nets to the Internet and WWW*, Oxford University Press, **2003**

16 Dehmer, M., Mowshowitz, A. *Information Sciences* 181 (**2011**), pp. 57

17 Holzinger, A., Dehmer, M., et al. *BMC Bioinformatics* 15 (**2014**), p. I1

18 Barabasi, A. L., Albert, R. *Science* 286 (**1999**), pp. 509

19 Kleinberg, J. *Nature* 406 (**2000**), pp. 845

20 Koontz, W., Narendra, P., et al. *IEEE Transactions on Computers* 100 (**1976**), pp. 936

21 Wittkop, T., Emig, D., et al. *Nature protocols* 6 (**2011**), pp. 285

22 Harary, F. *Structural models. An introduction to the theory of directed graphs*, Wiley, **1965**

23 Canutescu, A. A., Shelenkov, A. A., et al. *Protein science* 12 (**2003**), pp. 2001

24 Holzinger, A., Malle, B., et al. In *Interactive Knowledge Discovery and Data Mining: State-of-the-Art and Future Challenges in Biomedical Informatics, Springer LNCS 8401*, (eds. A. Holzinger, I. Jurisica), Springer, Berlin, Heidelberg, **2014**, pp. 57–80

25 Washio, T., Motoda, H. *ACM SIGKDD Explorations Newsletter* 5 (**2003**), p. 59

26 Jiang, C., Coenen, F., et al. *Knowledge-Based Systems* 23 (**2010**), pp. 302

27 Melcuk, I. *Dependency Syntax: Theory and Practice*, State University of New York Press, **1988**

28 Cook, D. J., Holder, L. B. *J. Artif. Int. Res.* 1 (**1994**), pp. 231

29 Yoshida, K., Motoda, H., et al. *Applied Intelligence* 4 (**1994**), pp. 297

30 Dehaspe, L., Toivonen, H. *Data Mining and Knowledge Discovery* 3 (**1999**), pp. 7

31 Fischer, I., Meinl, T. In *SMC (5)*, IEEE**2004**, pp. 4578–4582

32 Morales, L. P., Esteban, A. D., et al. In *Proceedings of the 3rd Textgraphs Workshop on Graph-Based Algorithms for Natural Language Processing*. TextGraphs-3, Association for Computational Linguistics, Stroudsburg, PA, USA, **2008**, pp. 53–56

33 Yan, X., Mehan, M. R., et al. *Bioinformatics* 23 (**2007**), pp. i577

34 Agirre, E., Soroa, A., et al. *Bioinformatics* 26 (**2010**), pp. 2889

35 Liu, H., Hunter, L., et al. *PLoS ONE*, 8

36 Liu, H., Komandur, R., et al. In *Proceedings of BioNLP Shared Task 2011 Workshop*, Association for Computational Linguistics, **2011**, pp. 164–172

37 Liu, J., Li, M., et al. *Pattern Recognition* 42 (**2009**), pp. 218

38 Rashevsky, N. *The bulletin of mathematical biophysics* 17 (**1955**), pp. 229

39 Trucco, E. *Bulletin of Mathematical Biology* 18 (**1956**), pp. 129

40 Mowshowitz, A. *The bulletin of mathematical biophysics* 30 (**1968**), pp. 533

41 Mowshowitz, A. *The bulletin of mathematical biophysics* 30 (**1968**), pp. 175

42 Körner, J. In *6th Prague Conference on Information Theory*. **1973**, pp. 411–425

43 Holzinger, A., Ofner, B., et al. In *Multidisciplinary Research and Practice for Information Systems, Springer LNCS 8127*, (Eds. A. Cuzzocrea, C. Kittl, D. E. Simos, E. Weippl, L. Xu), Springer, Heidelberg, Berlin, **2013**, pp. 354–362

44 Dehmer, M. *Symmetry* 3 (**2011**), pp. 767

45 Dehmer, M., Varmuza, K., et al. *Journal of chemical information and modeling* 49 (**2009**), pp. 1655

46 Shannon, C. E. *Bell system technical journal* 27 **1948**

47 Holzinger, A., Stocker, C., et al. *Entropy* 14 (**2012**), pp. 2324

48 Dehmer, M. *Appl. Math. Comput.* 201 (**2008**), pp. 82

49 Mowshowitz, A., Dehmer, M. *Entropy* 14 (**2012**), pp. 559

50 Dehmer, M., Mowshowitz, A. *Inf. Sci.* 181 (**2011**), pp. 57

51 Kessler, M. M. (1963). Bibliographic coupling between scientific papers. American documentation, 14(1), 10-25.

52 Kumar, S. *Aslib Journal of Information Management* 67 (**2015**) pp. 55

53 Culnan, M. J. (1987). Mapping the intellectual structure of MIS, 1980-1985: a co-citation analysis. Mis Quarterly, 341-353.

54 Small, H. (1973). CoâĂŘcitation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for information Science, 24(4), 265-269.

55 Hou, H., Kretschmer H., Liu, Z. *Scientometrics* 75 (**2008**), pp. 189