

Motif-based success scores in coauthorship networks are highly sensitive to author name disambiguation

David F. Klosik* and Stefan Bornholdt†

Institute for Theoretical Physics, University of Bremen, Hochschulring 18, 28359 Bremen, Germany

Marc-Thorsten Hütt‡

School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany

(Received 15 May 2014; published 19 September 2014)

Following the work of Krumov *et al.* [*Eur. Phys. J. B* **84**, 535 (2011)] we revisit the question whether the usage of large citation datasets allows for the quantitative assessment of social (by means of coauthorship of publications) influence on the progression of science. Applying a more comprehensive and well-curated dataset containing the publications in the journals of the American Physical Society during the whole 20th century we find that the measure chosen in the original study, a score based on small induced subgraphs, has to be used with caution, since the obtained results are highly sensitive to the exact implementation of the author disambiguation task.

DOI: [10.1103/PhysRevE.90.032811](https://doi.org/10.1103/PhysRevE.90.032811)

PACS number(s): 89.75.Hc, 89.65.-s, 01.30.-y

I. INTRODUCTION

Ever since the seminal work of Kuhn [1] it is widely accepted that the institutional process of knowledge production, i.e., science, cannot be fully described in purely logical, content-related terms, but has a significant social aspect to it. However, although the scientific community provides a comprehensive bookkeeping of its efforts by citing earlier work in new publications, only recently has this information been made widely accessible in the form of electronic datasets. With the aggregated citation information within a set of scientific publications at hand, one might now be able to quantitatively assess the extent to which the social embedding of science influences its structure and progression.

Traditionally, the focus of citation data analysis has been on the single publication level; indeed, the most prominent property of *paper citation networks* in which the vertices represent papers while directed edges represent the citations between the publications has been described by de Solla Price as early as 1965 [2]: The number of citations a paper receives (i.e., the corresponding node's in-degree) is broadly distributed, rendering highly cited publications significantly more frequent than they would be if scientists cited earlier work randomly. Similar broad degree distributions have been found in networks describing, e.g., technical, social, or biological interactions [3], and the question which might be the mutual underlying process, led to a now well-studied model class for network growth governed by a *rich-get-richer* effect, commonly referred to as preferential attachment [4–11]. In more recent times, citation data itself has mainly been used to quantify the assessment of scientific research thereby interpreting citations as indicators of impact or assignment of credit. To this end, numerous quantitative measures have been described which range from counting direct citations to considering also indirect citation paths [12–16] and some of

which aim at the scholar [17,18] or the journal level [19]. Additionally, there have also been structural investigations, e.g., regarding the community level [20] or other topological properties such as the richness of feed-forward loops in citation networks [21].

Another line of research has considered the collaboration network that can be constructed from citation data given the authorship metadata. More precisely, in a *coauthorship network* vertices represent authors and are connected by an undirected link if the two corresponding authors have coauthored one or more papers together. These networks have been investigated about a decade ago [22,23] with the main findings being the rather broad degree distribution and the strong small-world effect, i.e., the short paths between scholars in the network. There are also few approaches combining both citation and collaboration data, e.g., [24].

In this study we will construct a coauthorship network from a citation dataset, while the actual citation information is used to assess the success of the resulting links in the network.

While there have been studies about large-scale properties of coauthorship networks such as path lengths or community sizes and, of course, on the single-vertex scale, following Krumov *et al.* [25] here the focus will be on the intermediate level of small induced subgraphs known from the investigation of network motifs which originated in the biological context [26,27] and whose application has drawn some considerable criticism since the null-model graph ensemble has to be chosen with great care [28,29]. Note, however, that throughout this study *score* does not refer to the otherwise commonly used *z score*, since we are not interested in subgraph frequencies, but to the q_m value that is defined below. We will focus on the three- and four-node undirected induced subgraphs shown in Fig. 1.

We construct the collaboration network from a citation dataset provided by the American Physical Society (see Materials and Methods) by, first, identifying the articles' authors given in the article metadata. All author instances (i.e., names in author lists) are then grouped so that, ideally, all instances corresponding to the same actual scholar are represented by a single vertex in the network. There are different possible

*klosik@itp.uni-bremen.de

†bornholdt@itp.uni-bremen.de

‡m.huett@jacobs-university.de

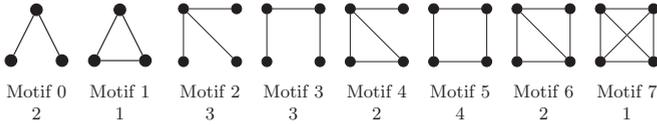


FIG. 1. The subgraphs used in this study. In the bottom line the minimum number of distinct papers required to build the subgraph is given.

implementations of this *author disambiguation task* which has become an object of investigation in its own right [30,31] with methods ranging from directly comparing the author names to incorporating metadata or even citation data. While in Krumov *et al.* one specific version has been used, here we find that the score proposed to assess the correlation between the collaboration pattern and the success of the corresponding papers is very sensitive to the exact implementation of the author disambiguation. We stress that this kind of ambiguity is not restricted to coauthorship networks but is also present in other network applications (see Sec. V). In the bibliography context, however, some effects of disambiguation errors have been pointed out in [32].

Furthermore, we show how the score is affected by the exclusion of large collaborations when the length of the author list above which papers are discarded is varied; in [25] this value was fixed at a number of eight.

II. COMPUTATION

First, we review the computation of the score proposed by Krumov *et al.* Two distinct authors who have coauthored are connected by an edge, e , that represents the list of all their mutually published papers, $P(e)$. Note that by this procedure a single publication can be represented by many edges. In order to assess whether there is a correlation between the collaboration structure on the small subgraph scale and the scientific impact of the publications contained in the subgraph's edges the number of citations to those papers is used to, first, compute the average number of citations to the papers of a single edge

$$w_2 = \langle c \rangle_e = \frac{1}{|P(e)|} \sum_{p \in P(e)} c(p) \quad (1)$$

and then take the average over all edges of all instances of a specific motif, M_m , i.e.,

$$q_m = \frac{1}{N_m} \sum_{m' \in M_m} q_{m,m'} = \frac{1}{N_m E_m} \sum_{m' \in M_m} \sum_{e \in m'} \langle c \rangle_e, \quad (2)$$

where $N_m = |M_m|$ gives the number of instances of the subgraph of type m and E_m yields the number of its edges (e.g., for the box motif $E_5 = 4$). A sketch of the procedure is given in Fig. 2. If one now shuffles the citation numbers among all publications the q_m is uniformly distributed, indicating that the motif-scale collaboration patterns in the given coauthorship network and the papers' success are not correlated after reshuffling. With the particular choice of (1) the motif scores are expected to yield the average number of citations of all papers in the network for every motif. The fact that this holds in all our computations implies that we obtain proper

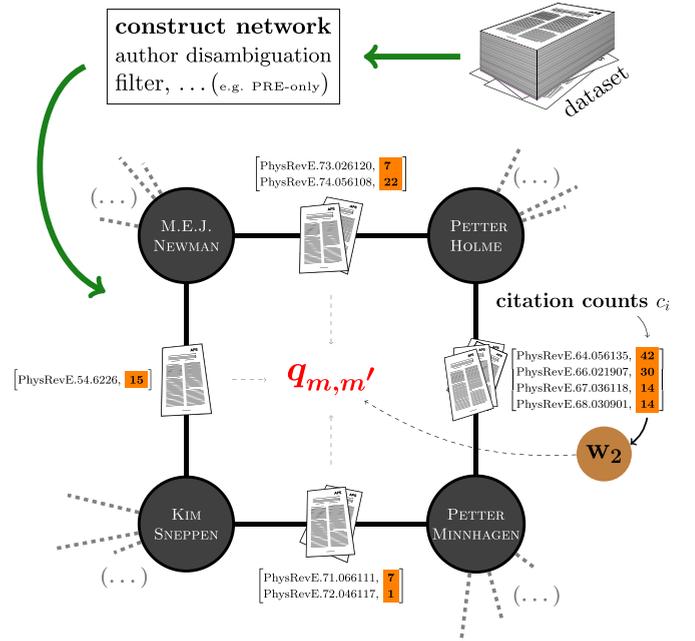


FIG. 2. (Color online) A sketch of the score computation.

averages of the often broadly distributed citation numbers. In [25] alternative edge scores have been proposed but with the current dataset (1) turns out to be the appropriate choice. It is important to notice that the shuffling procedure only affects the citation data on the edges of the otherwise fixed collaboration network, i.e., no topological shuffling is applied.

Krumov *et al.* report considerably higher q_m scores for the four-node motif called Motif 5 in Fig. 1 which they call the *box motif*. This subgraph stands out since it needs four distinct publications to be constructed (while, e.g., the four-node clique, Motif 7, might contain only one single mutually published paper) and there must not be any other collaborations between the four authors than in the author pairs corresponding to the four edges. The box motif therefore is an anticlustered structure and considerably fewer instances are found than of the other motif types. Again, we stress that we do not focus on motif frequency.

In order to keep the network topology and the dynamical quantity separated, in contrast to [25] we do not discard edges which are composed only of papers that did not receive any citation, but keep them in the network.

A. Maximum collaboration size

As mentioned above, in [25] all papers with more than eight authors have been discarded. Here we allow for different values of the *maximum collaboration size* (MCS) and investigate its influence on the q_m score (Sec. IV A). In terms of the social aspects encoded in a coauthorship network the exclusion of very large collaborations can be argued for since the thousands of links between their authors will hardly represent the same degree of personal connection as an edge between, e.g., two scholars publishing in a team of two.

B. Author name disambiguation

While one of the more elaborate author disambiguation schemes might have been used here, in order to assess the sensitivity of the score proposed in [25] we instead chose two versions of a very simple measure which consider only the authors' names as provided in the author lists of the publications in the dataset [33]. In the *all initials* (allINIT) disambiguation two author instance names are considered identical if in addition to the surname all initials are compatible (meaning that J. SMITH and JOHN SMITH are merged). The second implementation requires the full first name strings to match; we consider this *strict* (STRICT) since it will separate author instances if the first name is abbreviated in one and written out in the other case like the ones from the above example. However, there are surnames common enough to be shared by people with different given names; if these given names have compatible initials the allINIT method will spuriously merge them. In order to address this we apply a third disambiguation which we will refer to as SPLIT: We apply independently the allINIT and STRICT disambiguation to the same data and then track those allINIT names that are split into more than two distinct authors in the STRICT implementation. All papers with those names among the authors are then filtered. On the resulting dataset we then perform all remaining steps of the calculation with the allINIT method applied. Note that the result of this procedure depends on all previous data filtering, especially the one according to the MCS, so the number of remaining papers has to be compared to the number of publications in the original data after discarding those with more than MCS authors. In the cases shown here the data had been prepared with a MCS value of 10 before the application of the SPLIT filtering and we find that while for the merge of the journals 36% and in PRC 32% are filtered, in the remaining single journal portions around 80% of the papers are kept. For comparison, if one excluded all papers from the dataset that had at least one author whose given name is only provided in initialized form, not only would over half of the dataset be excluded, but also this filtering would be biased against large collaborations since there name abbreviation is used to save printing space.

III. MATERIALS AND METHODS

The citation data used in this study is composed of all APS journals published between July 1893 and December 2009 and the associated information about citations between these papers, and may therefore be assumed to cover many major 20th century contributions to physics. The dataset can be requested at the American Physical Society [34]. The data is provided in separate files corresponding to the individual APS journals which emphasize different topics; we chose to exclude *Reviews of Modern Physics* and the *online-only* journals *Physical Review Special Topics - Accelerators and Beams* and *Physics Education Research* due to their considerably smaller size compared to the remaining journals. These remaining journals vary in size from *Physical Review E* with 35 022 to *Physical Review B* with 133 269 publications. In addition to extensive bibliographic metadata such as authorship, publication history, and PACS numbers the dataset also provides *article type* tags,

thereby allowing us to filter nonstandard material (i.e., those tagged *comment*, *erratum*, *reply*, *editorial*, *essay*, *publisher note*, *retraction*, *miscellaneous*) and restrict our analysis to standard research publications.

Since motif enumeration is computationally costly, for most of the calculations the RANDESU algorithm as described by Wernicke [35] is applied which instead of enumerating all subgraphs the algorithm samples uniformly from the set of all motifs. By performing duplicate runs of the motif score computation with the same parameters we checked that the scores are not sensitive to the sampling procedure. For smaller networks the full enumeration is performed by application of the ESU algorithm, presented also in [35].

IV. RESULTS

We computed the q_m score in both the single journal portions and the whole APS dataset and found that the shuffling of the citation frequencies among the papers indeed yields a uniform distribution of the scores. As mentioned above, in all cases the appropriate edge weight to achieve this was $w_2 = \langle c \rangle_e$, i.e., the averaged citation frequency of all papers on the corresponding edge. In the following we show the influence of a variation of the maximal number of authors above which a paper is excluded from the data, as well as the sensitivity of the q_m score against different implementations of the author disambiguation task.

A. MCS scan

In Fig. 3 the average network degree as a function of the MCS value is shown for both the allINIT and the STRICT disambiguation. Increasing the MCS value translates into the introduction of new potential nodes and edges to the network, and since a publication with a authors can produce up to $a(a-1)/2$ new edges $\langle k \rangle$ grows. Unlike PRA, PRB, and PRE, the journals PRC, PRD, and PRL which feature larger collaborations do not show a saturated average degree in the shown MCS interval. The fact that the STRICT disambiguation results in more distinct authors than the allINIT implementation is reflected by the lower average degrees.

Especially with the STRICT disambiguation scheme we can reproduce the main finding of [25], the higher q_m score for the box motif, for small enough values of the maximum collaboration size. Although in few journal portions this result

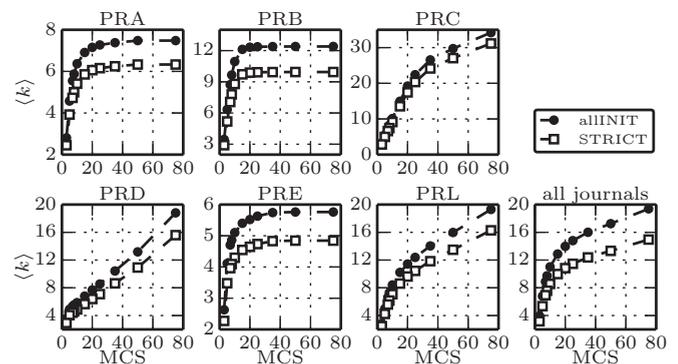


FIG. 3. A scan of the MCS value translates to a variation of the average degree.

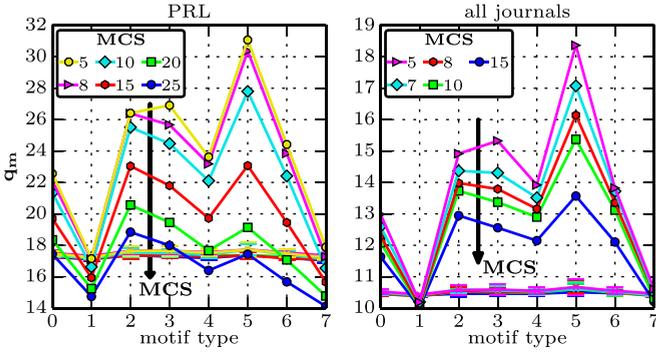


FIG. 4. (Color online) Two example MCS scans (STRICT disambiguation). The box motif signal is weaker for larger MCS. The almost flat lines correspond to the averaged results of 30 runs with shuffled citation counts.

is rather robust, in most cases the box motif signal tends to decrease with increasing MCS; Fig. 4 shows two examples: In the merge of the journals the box motif keeps the highest score in the given MCS range, while in PRL the box motif signal not only gets less pronounced but is lost at $MCS \approx 15$.

B. Influence of the author disambiguation

The exact implementation of the author disambiguation turned out to be crucial considering the q_m distribution. From the network perspective it can be interpreted as a local perturbation (as illustrated in the sketch in Fig. 5) which strongly influences the q_m scores computed on the few-node subgraph scale. While the SPLIT disambiguation shows a rather similar behavior to the STRICT case (only PRA and PRL do not show a maximum score for the box motif in the former case), by switching from the STRICT to the allINIT disambiguation scheme the q_m score distribution can change qualitatively. For example, the journal merge does not show the box motif signal any longer and also in the case of PRB and PRE shown in Fig. 5 the box motif scores are suppressed.

C. Distributions of the $q_{m,m'}$ scores

The q_m are averaged values over the $q_{m,m'}$ scores of the single instances of the specific motif type m and as such are influenced of course not only by the top-ranked but also by the motif instances with lowest scores. In Fig. 6 the distributions of the $q_{m,m'}$ values for the different motif types are shown for the example of the PRB network portion. In the depicted range the distributions can be grouped according to the minimal number of distinct papers required to construct the specific motif (bottom row in Fig. 1). In the box motif which requires the largest minimum number of distinct papers very low $q_{m,m'}$ scores are suppressed while the three- and four-node cliques which need only one mutual publication among the authors often show very small scores. This is due to the distribution of the citation counts used in the computation of the scores: The many poorly cited publications can translate directly to low scores for cliques, while it is unlikely for the at least four edges of a box motif to exclusively consist of poorly cited papers. Indeed, the $q_{m,m'}$ distributions of the motif types 0, 4, 6 and 2, 3 which require the intermediate number of 2 and 3

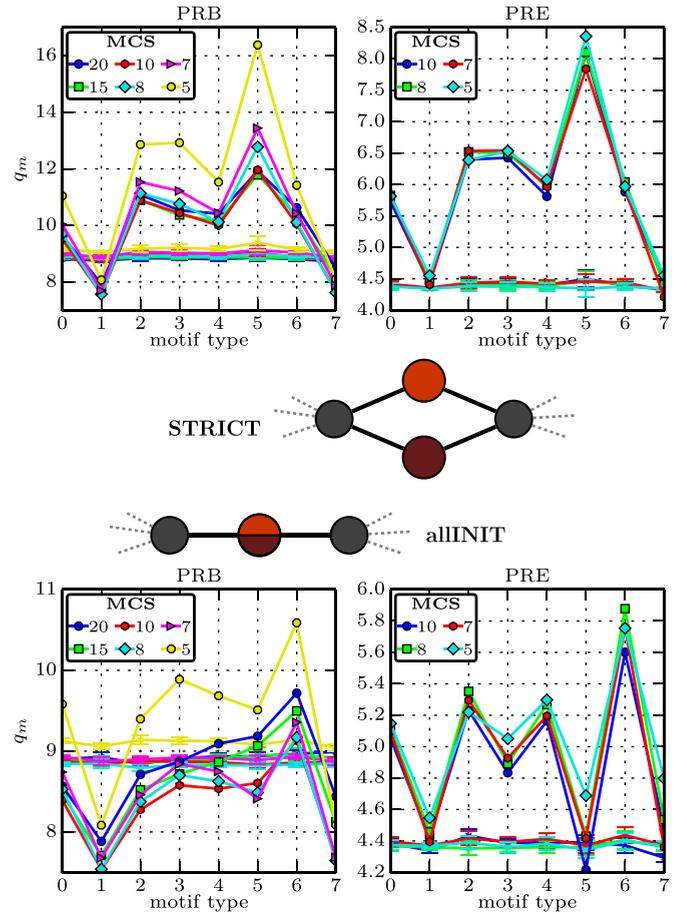


FIG. 5. (Color online) The influence of the local topological perturbation introduced by switching between the STRICT (top) and the allINIT (bottom) disambiguation scheme for the PRB and the PRE single journal portions. Again, the almost flat lines correspond to the averaged results of 30 runs with shuffled citation counts.

edges, respectively, is consistent with this interpretation. This renders the investigation of the top motif instances according to the $q_{m,m'}$ scores less promising.

A paper with a authors contributes to $a(a-1)/2$ edges which in turn may contribute to very many motif instances. One corresponding effect can be seen when examining the top motif instances according to the $q_{m,m'}$ scores: These lists can be dominated by highly cited publications. For example, more than half of the top motifs in the PRA journal portion (SPLIT disambiguation, $MCS = 10$) share the publication “*Quantum computation with quantum dots*” by DiVincenzo and Loss [36]. If one is not, however, interested in compiling lists of top instances, but only in averaged values like the q_m , another approach might be used. The q_m for a specific motif type m can be rephrased as a weighted sum over the citation counts of all papers that contribute to instances of that motif; the weights, however, depend not only on how often the paper contributes but also on the numbers of other papers it shares its edges with:

$$q_m = \sum_{i \in P} c_i \frac{1}{N_m E_m} \sum_{m' \in M_m} \sum_{e \in m'} \frac{1}{|P^{(i)}(e)|},$$

$$|P^{(i)}(e)| = \infty \quad \text{if } i \notin P(e) \quad (3)$$

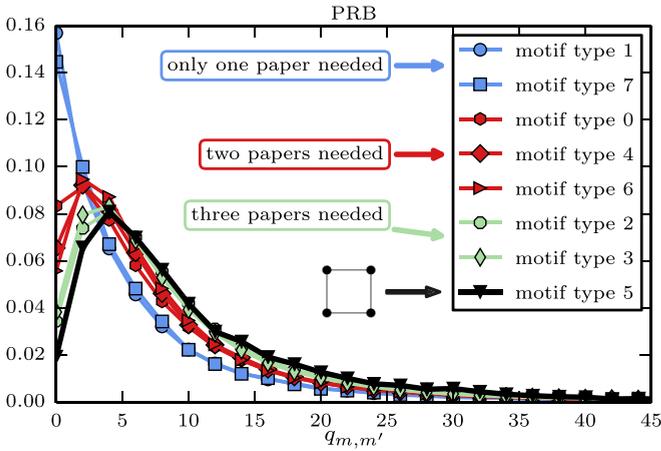


FIG. 6. (Color online) Normalized frequency of the $q_{m,m'}$ scores for the different motif types. The PRB journal portion has been used with $MCS = 10$ and the SPLIT filtering has been applied.

with P being the set of all papers. A simple alternative would be to set these weights to $\delta_{i \in M_m} / |P_m|$ where P_m is the set of all papers that at least once contribute to a subgraph instance of type m , i.e., to compute the average citation frequency of the papers that contribute to the motif type in question but consider each publication only once.

Indeed, the box motif shows the highest average citation counts compared to the other motif types for all single journals as well as the merge of all journals independently of the choice of the disambiguation scheme and for MCS values not smaller than 5. In Fig. 7 these citation averages are given for the example of the PRB and the PRE data and an MCS value of 10 (compare to Fig. 5). In the figure a box plot of the

average of a random sample of size $|M_m|$ drawn from the citation distribution is given, showing that the average citation count for the box motif lies significantly above the sample average and is therefore not just due to the smaller sample size. However, since uniform sampling assumes that every paper is a part of every motif type with the same probability, the sampling procedure might be refined to incorporate the specific conditions imposed by the different motif types. In a first estimation, we restricted the samples for the box motif to those papers that have at least two coauthors with at least degree 2 each (i.e., have at least one other collaborator) for the case of the SPLIT disambiguation. Although this increases the sample average, the measured average citation count is still several (at least 5) standard deviations above the sample average.

V. DISCUSSION AND CONCLUSION

Motivated by the question to what extent small-scale patterns of collaboration among scholars influence the recognition of the resulting papers within the scientific community (in terms of the number of citations the publications receive) we searched induced few-node subgraphs (known from motif analysis) in a comprehensive citation dataset provided by the American Physical Society. We then assigned a score to the detected subgraphs as described in Krumov *et al.* [25] based on the citation counts of the corresponding publications so that an average value, q_m , for every subgraph type m can be computed. It turns out that this score is sensitive to certain details of the network aggregation from the data and the main result of the aforementioned study [25], i.e., the highest score for the anticlustered box motif, cannot be seen clearly in all cases: For example, discarding different amounts of the largest collaborations which translates to a variation of the average network degree influences the q_m signature. A stronger effect can be seen when the network is locally perturbed by changing the implementation of the author disambiguation task; in the *all initial* method of author disambiguation the box-motif score can be suppressed, while a stricter disambiguation scheme yields a high score. Consequently, such a score should be used with care. This result might be of interest in other applications of few-node induced subgraph-based scores to data which is incomplete or not unambiguously processed into a graph representation since it highlights a possible difficulty on the motif scale in addition to the well-known problems concerning the proper null-model selection in the usual motif-analysis. While the ambiguity concerning whether two author names are to be translated into one or two nodes is exclusive to coauthorship networks, several other types of complex networks have related underlying disambiguation tasks: Enzyme-centric metabolic networks (where the nodes are enzymes and a link represents common substrates and products) can also be formulated on the level of reactions (see, e.g., [29]); for many organisms, the enzyme inventory of the genome is incomplete [37,38] and the enzyme-to-reaction mapping introduces similar issues as the author disambiguation. On the basis of the present investigation we expect that the topological properties of enzyme-centric and reaction-centric metabolic networks show similar differences as the coauthorship networks under disambiguation schemes

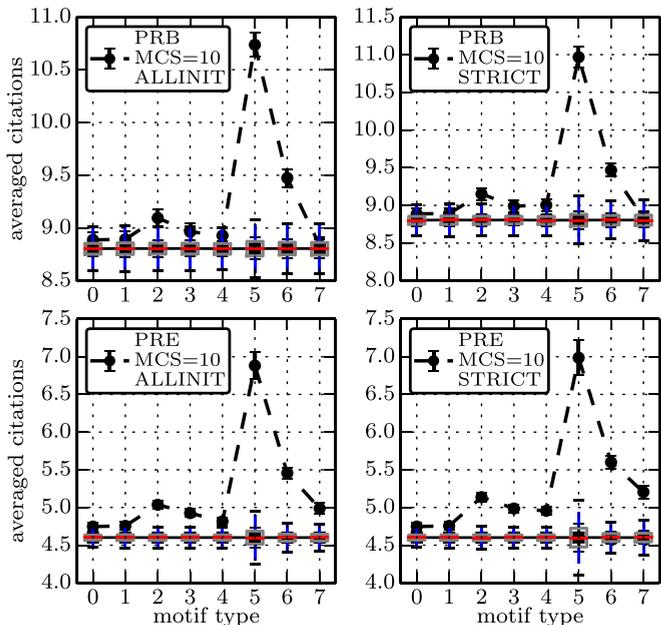


FIG. 7. (Color online) The average citation counts of the papers that contribute to at least one instance of a given motif type for the example of the PRB and PRE data. Compare to Fig. 5 where also the allINIT and the STRICT disambiguation is shown.

of different stringency. For bacteria, transcriptional regulatory networks (where the nodes are genes and a directed link from gene g_1 to gene g_2 represents the regulation of a g_2 by a transcription factor produced by g_1) are sometimes more adequately described on the level of groups of genes, called *operons*, under common regulation (see, e.g., [39]). In eukaryotic organisms, attempts to apply this concept lead to ambiguities [40], similar to those encountered in the author disambiguation. In computational neuroscience, the topology of cortical area networks is often analyzed on the basis of “regions of interest” (ROI) from diffusion spectrum imaging,

rather than functionally defined cortical areas [41]. Relating these ROI to the functional cortical areas, and comparing results obtained with different segmentations of the cerebral cortex, again, is comparable to the author disambiguation task.

ACKNOWLEDGMENT

We acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) under Contracts No. BO 1242/6-1 and No. HU 937/9-1.

-
- [1] T. S. Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, Chicago, 1996).
- [2] D. J. D. S. Price, *Science* **149**, 510 (1965).
- [3] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [4] D. J. D. S. Price, *J. Am. Soc. Inf. Sci.* **27**, 292 (1976).
- [5] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [6] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. Lett.* **85**, 4633 (2000).
- [7] S. N. Dorogovtsev and J. F. F. Mendes, *Phys. Rev. E* **63**, 056125 (2001).
- [8] S. N. Dorogovtsev and J. F. F. Mendes, *Phys. Rev. E* **62**, 1842 (2000).
- [9] G. Bianconi and A.-L. Barabási, *Phys. Rev. Lett.* **86**, 5632 (2001).
- [10] P. L. Krapivsky and S. Redner, *Phys. Rev. E* **63**, 066123 (2001).
- [11] P. L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).
- [12] F. Radicchi, S. Fortunato, and C. Castellano, *Proc. Natl. Acad. Sci. USA* **105**, 17268 (2008).
- [13] J. Bollen, H. Van de Sompel, A. Hagberg, and R. Chute, *PLoS ONE* **4**, e6022 (2009).
- [14] P. Chen, H. Xie, S. Maslov, and S. Redner, *J. Informetrics* **1**, 8 (2007).
- [15] D. Walker, H. Xie, K.-K. Yan, and S. Maslov, *J. Stat. Mech.: Theory Exp.* (2007) P06010.
- [16] D. F. Klosik and S. Bornholdt, [arXiv:1301.7471](https://arxiv.org/abs/1301.7471).
- [17] J. E. Hirsch, *Proc. Natl. Acad. Sci. USA* **102**, 16569 (2005).
- [18] F. Radicchi, S. Fortunato, B. Markines, and A. Vespignani, *Phys. Rev. E* **80**, 056103 (2009).
- [19] J. D. West, T. C. Bergstrom, and C. T. Bergstrom, *Coll. Res Libr.* **71**, 236 (2010).
- [20] P. Chen and S. Redner, *J. Informetrics* **4**, 278 (2010).
- [21] Z.-X. Wu and P. Holme, *Phys. Rev. E* **80**, 037101 (2009).
- [22] M. E. J. Newman, *Phys. Rev. E* **64**, 016131 (2001).
- [23] M. E. J. Newman, *Phys. Rev. E* **64**, 016132 (2001).
- [24] T. Martin, B. Ball, B. Karrer, and M. E. J. Newman, *Phys. Rev. E* **88**, 012814 (2013).
- [25] L. Krumov, C. Fretter, M. Müller-Hannemann, K. Weihe, and M.-T. Hütt, *Eur. Phys. J. B* **84**, 535 (2011).
- [26] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
- [27] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, *Science* **303**, 1538 (2004).
- [28] C. Fretter, M. Müller-Hannemann, and M.-T. Hütt, *Phys. Rev. E* **85**, 056119 (2012).
- [29] M. E. Beber, C. Fretter, S. Jain, N. Sonnenschein, M. Müller-Hannemann, and M.-T. Hütt, *J. R. Soc., Interface* **9**, 3426 (2012).
- [30] C. Schulz, A. Mazloumian, A. M. Petersen, O. Penner, and D. Helbing, [arXiv:1401.6157](https://arxiv.org/abs/1401.6157).
- [31] T. C. Silva and D. R. Amancio, *Chaos* **23**, 013139 (2013).
- [32] B. D. Fegley and V. I. Torvik, *PLoS ONE* **8**, e70299 (2013).
- [33] S. Milojević, *J. Informetrics* **7**, 767 (2013).
- [34] See <https://publish.aps.org/datasets>.
- [35] S. Wernicke, *IEEE/ACM Trans. Comput. Biol. Bioinf.* **3**, 347 (2006).
- [36] D. Loss and D. P. DiVincenzo, *Phys. Rev. A* **57**, 120 (1998).
- [37] A. M. Feist, M. J. Herrgård, I. Thiele, J. L. Reed, and B. O. Palsson, *Nat. Rev. Microbiol.* **7**, 129 (2009).
- [38] J. Schellenberger, J. O. Park, T. M. Conrad, and B. O. Palsson, *BMC Bioinf.* **11**, 213 (2010).
- [39] C. Marr, M. Geertz, M.-T. Hütt, and G. Muskhelishvili, *BMC Syst. Biol.* **2**, 18 (2008).
- [40] T. Blumenthal, *Briefings Funct. Genomics Proteomics* **3**, 199 (2004).
- [41] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, *PLoS Biol.* **6**, e159 (2008).