# Coauthorship networks and patterns of scientific collaboration

M. E. J. Newman

*Center for the Study of Complex Systems and Department of Physics, University of Michigan, Ann Arbor, MI 48109*

**Abstract**

Using data from three bibliographic databases in biology, physics, and mathematics respectively, networks are constructed in which the nodes are scientists and two scientists are connected if they have coauthored a paper together. We use these networks to answer a broad variety of questions about collaboration patterns, such as the numbers of papers authors write, how many people they write them with, what the typical distance between scientists is through the network, and how patterns of collaboration vary between subjects and over time. We also summarize a number of recent results by other authors on coauthorship patterns.

## 1 Introduction

It has long been realized that the coauthorship of articles in learned journals provides a window on patterns of collaboration within the academic community. Coauthorship of a paper can be thought of as documenting a collaboration between two or more authors, and these collaborations form a "coauthorship network" such as that depicted in Fig. 1, in which the network nodes represent authors and two authors are connected by a line if they have coauthored one or more papers. The structure of such networks turns out to reveal many interesting features of academic communities.

Networks are not new to bibliometrics. The field has a long history of the study of citation networks [1, 2], the networks formed by the citations between papers. These are quite distinct, however, from coauthorship networks—the nodes in a citation network are papers, not authors, and the links between them are citations, not coauthorship. The coauthorship network is as much a network depicting academic society as it is a network depicting the structure of our knowledge. And, perhaps because of this, it has received far less attention than have citation networks. Nonetheless, it has much of value to tell us, as recent work has shown.

During the 1990s (and possibly earlier) a number of authors pointed out the potential utility of coauthorship data, and in some cases performed small-scale

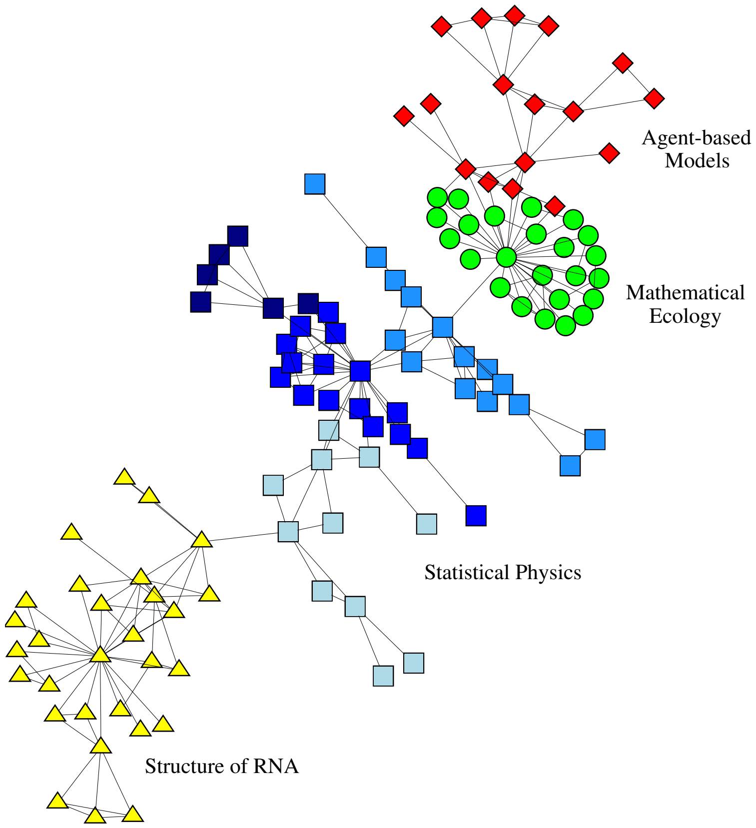


Figure 1: An example of a small coauthorship network depicting collaborations between scientists at a private research institution. The network appears to divide into a number of subcommunities, as indicated by the shapes of the nodes, and these subcommunities correspond roughly to topics of research, as indicated. After Girvan and Newman [3].

statistical analyses of such things as frequency of coauthored articles by particular authors or authors at particular institutions [4–8]. But it was with the advent of comprehensive online bibliographies that construction of complete or near-complete coauthorship networks for entire fields became a realistic possibility. Starting around 2000 several researchers began the construction of large-scale networks representing research in mathematics [9–11], biology, physics and computer science [12] and neuroscience [11].

In this paper we look in detail at three particular networks of scientific collaborations and describe some of the patterns they reveal. The networks are:

1. A network of coauthorships of papers in the Medline bibliographical database from 1995 to 1999 inclusive. Medline is a widely used and compendious database of papers covering biomedical research. Biomedical research accounts for the largest part of civilian scientific research by far, dwarfing research in all other subjects put together in terms of expenditure. Any study that excluded biomedicine could not claim to be representative of science as it is practiced today.

2. A network of coauthorships of physicists assembled from papers posted on the widely used Physics E-print Archive at Cornell University (formerly at the Los Alamos National Laboratory) between 1995 and 1999. Physics has led the way in moving from journal publication to author self-publication in online preprint databases, with preprint publication largely replacing journal publication in some subfields. Preprint databases provide a useful source of up-to-the-minute publication records, although their coverage is less complete than that of professionally maintained databases like Medline.

3. A collaboration network of mathematicians compiled from databases maintained by the journal *Mathematical Reviews*. Of the networks yet studied, this is probably the most complete and accurate, covering the period from 1940 to the present without any break.

Networks 1 and 2 where constructed by the present author from bibliographic data supplied by the maintainers of the corresponding databases. Network 3 was constructed by Jerrold Grossman and Patrick Ion [9] and graciously supplied to the author by Prof. Grossman.

In Section 2 we describe a variety of results derived from analysis of these networks, and highlight some differences between the three subjects studied. In Section 3 we summarize some recent additional results obtained using the same or additional data, including a number of results due to other authors. In Section 4 we give our conclusions. Many of the results reported here have appeared previously in Refs. 12–16, as well as a number of other papers, which are cited as appropriate.

| | biology | physics | mathematics |
|---|---|---|---|
| number of authors | 1520251 | 52909 | 253339 |
| number of papers | 2163923 | 98502 | – |
| papers per author | 6.4 | 5.1 | 6.9 |
| authors per paper | 3.75 | 2.53 | 1.45 |
| average collaborators | 18.1 | 9.7 | 3.9 |
| largest component | 92% | 85% | 82% |
| mean distance | 4.6 | 5.9 | 7.6 |
| largest distance | 24 | 20 | 27 |
| clustering coefficient | 0.066 | 0.43 | 0.15 |
| assortativity | 0.13 | 0.36 | 0.12 |

Table 1: Summary statistics for the three coauthorship networks analyzed here. After Newman [13] and Grossman [10].

# 2 Statistical properties of coauthorship networks

A summary of the basic statistics of the three networks studied here is given in Table 1. The largest of the networks is, not surprisingly, the biomedical network, with more than 1.5 million authors over a five year period. Even the mathematics network, which covers a much longer period (about sixty years), comes nowhere close to this size. Clearly biology dwarfs other subjects in terms not only of spending but also of manpower. The number of papers shows a similar pattern, although we do not have precise data for the number of papers in the mathematics database. (Grossman [10] cites a figure of "about 1.6 million authored items" in the *Mathematical Reviews* database, for a slightly more recent version of the network than that studied here.)

The number of papers per author is similar across the three subject areas, lying between five and seven in each case. Since the mathematics database covers a longer time period however, this may indicate that mathematicians are producing fewer papers than their more empirically minded colleagues in the sciences. Scientific productivity, measured by number of papers authored, has had a long history of study in bibliometrics with the articles by Lotka [17] and Shockley [18] being famous early examples. Both these authors found that the number of papers produced by scientists had a "fat-tailed" distribution in which a small number of scientists produced a very large number of papers, a result which has since been confirmed by others [19, 20], and which is seen in our own data as well [12, 13].

The number of authors per paper, by contrast, varies substantially between the subjects studied, with biology having the largest number and mathematics having the smallest. This presumably reflects real differences in the way research is done in these fields, with biological research consisting often of work by large groups of laboratory scientists, and mathematics consisting of theoretical work done primarily by individuals alone or by pairs of collaborators.

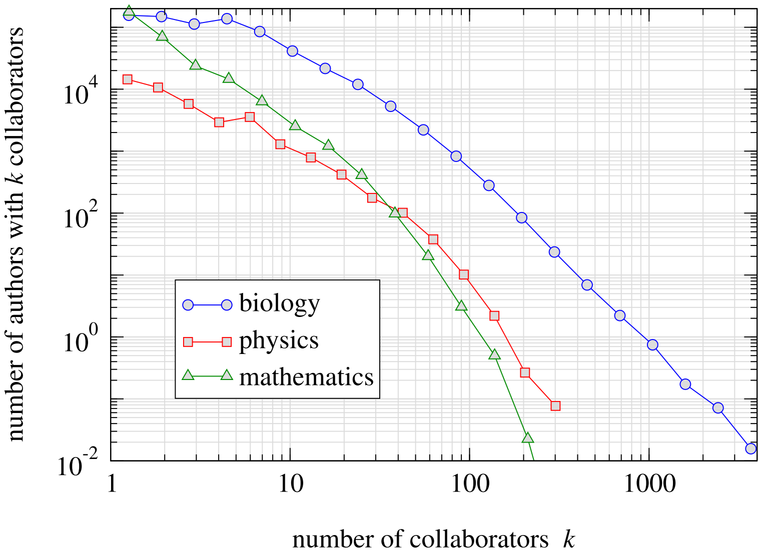


Figure 2: Histograms of the distribution of numbers of collaborators for scientists in each of three fields studied.

Grossman [10] says that 66% of mathematics papers are written by a single author alone (although this number changes over time—see Section 3). In the Medline database, the corresponding figure is 21%. These figures may offer some explanation for the possible lower productivity of mathematics in terms of papers published per unit time: with fewer coauthors on most publications the production of a mathematics paper involves more work per author.

A similar pattern is revealed in the average number of collaborators an individual has in the three fields, which is more than four times higher in biology than in mathematics. This again is presumably a result of different modes of research, with biology being primarily experimental, mathematics being entirely theoretical, and physics being a combination of the two. (The quintessential example of scientific experiment on an industrial scale is high-energy physics, for which it was found, using the SPIRES database of high-energy physics papers, that authors had an amazing 173 collaborators on average over the five-year period from 1995 to 1999 [12].)

In addition to mean numbers of papers and coauthors, one can look at the distributions of these quantities. In Fig. 2, for instance, we show the distribution of the number of coauthors that scientists have for the three subjects. As the figure shows, the distributions are quite similar for the three subjects, although the distribution for biomedicine (circles) has a longer tail, reflecting the higher mean number of collaborators that individuals have in that field. In each case the distribution is fat-tailed—like the distribution of number of papers written by scientists mentioned above—with a small fraction of scientists having a very large number of collaborators, up to thousands in the case of the biology network. (Recall that the data for this network cover only a five-year period; publishing papers with one thousand coauthors in less than two thousand days is an impressive achievement by any measure.) Unlike some other networks, such as the World Wide Web and the Internet, however, the distributions for these networks do not follow power laws—they are not "scale-free networks" in the jargon of the field. It has been suggested that the distributions are actually power-law in form with an exponential cutoff [10, 12, 21], and this appears to be a reasonable fit to the data. The cutoff may be produced by the finite time window used in the study, a hypothesis that could in principle be tested by varying the size of that window, although we do not do this here.

Table 1 also gives the size of the largest component in each of the networks. A component is a set of network nodes connected via coauthorship such that any node in the set can be reached from any other by traversing a suitable path of intermediate collaborators. For each of the networks studied here the largest component fills most of the network—occupying between 82% and 92% of the network in the three cases. Thus a large portion of each of these communities is connected together in a kind of linked research enterprise, rather than working separately in isolation. Overall this seems a promising picture—intellectual isolation from the mainstream of one's research area cannot often be a good thing. Most of the scientists who do not belong to the largest component are members of small disconnected components containing only a handful of others at most.

Many recent studies of networks of various types have focused on network distance between nodes. This distance is define as the number of "hops" along links in the network that one needs to make to move from one given node to another. A pair of individuals who have coauthored a paper, for instance, are distance 1 apart, while a pair who have not but who share a common coauthor are distance 2 apart, and so forth. In the late 1950s, Kochen and Pool [22] speculated on mathematical grounds that networks might show surprisingly small typical distances between pairs of nodes, and in a famous experiment some years later, Milgram demonstrated that this was the case for acquaintance networks, at least in the United States [23, 24]. Our coauthorship networks appear to follow the same pattern. We calculate the distance between all pairs of individuals in a network using a breadth-first search or "burning" algorithm [25], and then take the average to give the figures shown in the table. For each of the networks the result is very small, at least compared with the size of the network. Mathematics has the largest mean distance, possibly again a result of the relative sparsity of mathematics collaborations, but even its value of 7.6 is tiny compared with the quarter of a million mathematicians in the network. This appears to indicate a close-knit cohesive network in which most people are not only connected to others by some path through the network, but are connected by a short one.

A certain amount of attention has focused also on the distances from *particular* scientists to others in the coauthorship network. Mathematicians have long discussed the "Erdős number," which is the distance through the mathematics

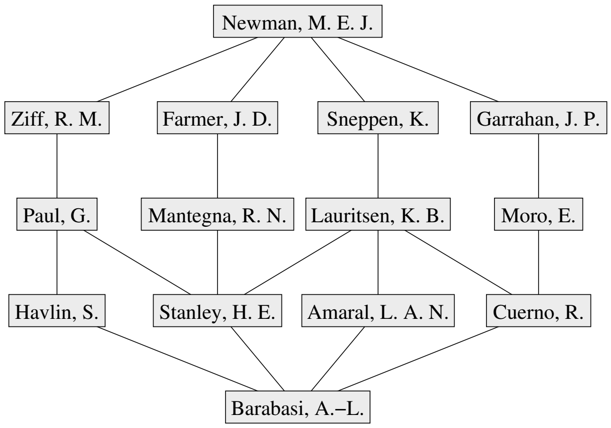


Figure 3: The shortest paths through the collaboration network of physics papers from the present author to László Barabási, who also publishes on networks.

network from a given mathematician to Paul Erdős, an influential Hungarian number theorist of the twentieth century who was renowned for his prolific publication and collaboration. Erdős numbers have been studied in depth by a number of authors using the *Mathematical Reviews* data [9, 10, 26, 27]. It is found, for instance that the mean distance from Paul Erdős to other mathematicians is much lower than the mean distance in the network as a whole, taking a value around 4.7. (Similar mean distances from other individuals, most of which are significantly higher than Erdős's mean, are sometimes called "Doe numbers" [10].) The largest distance or "diameter" of a network is also occasionally of interest. It is between 20 and 30 for each of the networks studied here.

It is straightforward to create a computer algorithm to find the shortest path between two particular scientists, again using breadth-first search, and it has been suggested by Kautz *et al.* [28] that such programs could be of use for providing "referral chains"—links of acquaintances that individuals could use to establish contact with other scientists. In the simplest case, for example, it might be useful to know that one shared a common collaborator with another scientist if one wished to arrange an introduction. Note that the shortest path between two individuals need not be unique, and in fact it happens quite frequently that there are two or more shortest paths of equal length. Fig. 3 shows shortest paths in the network of of the Physics E-print Archive between the present author and László Barabási of the University of Notre Dame, who also publishes on networks. As the figure shows, there are several different paths that go from the one scientist to the other, all with length four. This particular case is interesting because it shows that scientists working in the same field need not be linked through others in their field. The shortest paths in this case

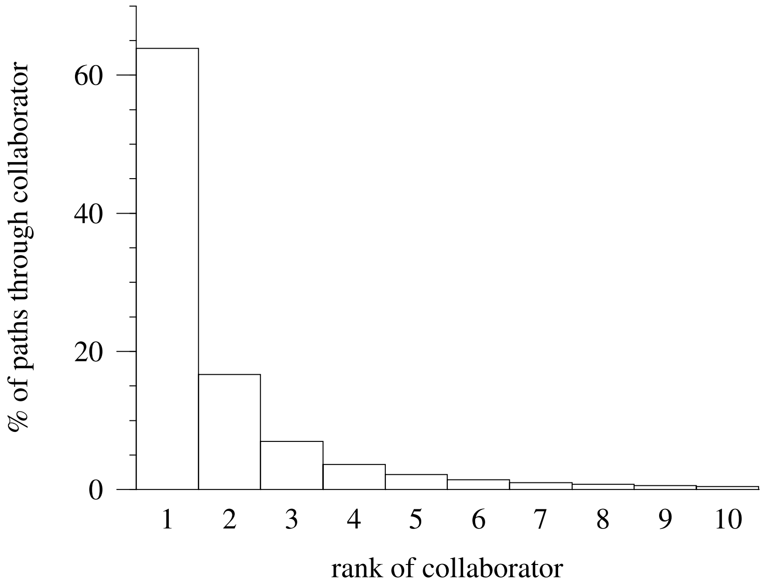


Figure 4: The average percentage of paths from other scientists to a given scientist that pass through each collaborator of that scientist, ranked in decreasing order. The plot is for the physics network, although similar results are found for the others.

are established via the collaborations of the present author with Doyne Farmer, Juan Pedro Garrahan, Kim Sneppen, and Robert Ziff, only the last of which collaborations involved work on networks (and then only peripherally).

Another interesting network measure related to shortest paths has been suggested by S. H. Strogatz, who asks how many of the shortest paths from a particular individual to others pass through each of their collaborators. Is it the case that most of your connections to others are via just one or two of your best-connected collaborators, or are they distributed evenly among your collaborators? For the networks studied here it turns out that the former is the case, as shown in Fig. 4. This figure shows for the physics network what percentage of shortest paths pass through each of a scientist's coauthors on average. Thus the figure reveals that on average about 64% of an individual's shortest paths to other pass through the best-connected of their collaborators, and most of the remainder pass through the next-best-connected. This appears to indicate that a small number of scientists are playing the role of broker for communications between others. (See also the discussion of betweenness centrality in Section 3.)

Two other quantities of interest, both previously studied for many networks, are also given in Table 1. The first is the "clustering coefficient" [29], which measures network "clustering" or "transitivity"—the probability that two of a scientist's coauthors have themselves coauthored a paper together. In topological terms it is a measure of the density of triangles in a network, a triangle being formed every time two of your collaborators collaborate also with each other. The clustering coefficient is highest for physics (43%) and lowest for biology (7%), and it is unclear why there is so much variation between fields. Presumably the figures reflect substantial differences between collaboration patterns in the sciences, but what these differences are is far from obvious. Part of the clustering in each network can be accounted for by papers with three or more coauthors. Such papers introduce triangles of collaborating authors and hence increase the clustering coefficient. This effect can only account for about a half of the clustering seen in coauthorship networks however [30]; the rest must be due to sociological or organization effects of some kind.

An alternative way to measure the clustering effect is to look at the time evolution of the network. Among social networks, coauthorship networks are unusual in having well-documented time evolution. Since each paper comes with a date of publication, we can say when each connection was added to the network and so we can reconstruct the order in which the network grew. This allows us to ask what the probability is of two scientists coauthoring a paper, given that they have a third mutual collaborator and that they haven't collaborated in the past. By studying only scientists who have not previously collaborated, we eliminate any bias introduced by papers with three or more coauthors. In 15 we showed that scientists with a single mutual collaborator are about 45 times more likely to coauthor a paper than those with no mutual collaborator. Those with two are more than 100 times as likely to coauthor a paper.

The last line of the table gives the "assortativity coefficient" or degree correlation for the networks [16]. This is the correlation coefficient for the number of collaborators that coauthors have. It lies in the range from $-1$ to 1, with positive values indicating that people with many collaborators tend to collaborate with others with many collaborators and negative values indicating the reverse. The coefficient is positive for all the networks studied, indicating that the most gregarious scientists tend to be connected to each other. Again it is an open question why this should be the case.

It is also possible to extract from coauthorship data a measure of the strength of the collaboration between pairs of individuals. The simplest such measure would just be a count of the frequency with which two scientists have coauthored papers—the number of coauthored papers over a given interval, for instance. However, this fails to take into account the number of other coauthors on each paper. Presumably two authors who collaborate on a ten-author paper are, in general, working less closely with one another that two who produce a two-author paper with no other help. To account for this effect, we proposed in Ref. 14 the measure of collaboration strength illustrated in Fig. 5. Each paper coauthored by a given author pair adds an amount $1/(n-1)$ to the strength of their collaboration, where $n$ is the total number of authors on the paper. The rationale behind this choice is that an author divides his or her time between the $n-1$ other authors with whom he or she works on a paper, and hence the strength of the connection to each of them varies inversely as $n-1$. As an example of this measure, we show in Fig. 6 the coauthors of Gerard Barkema—

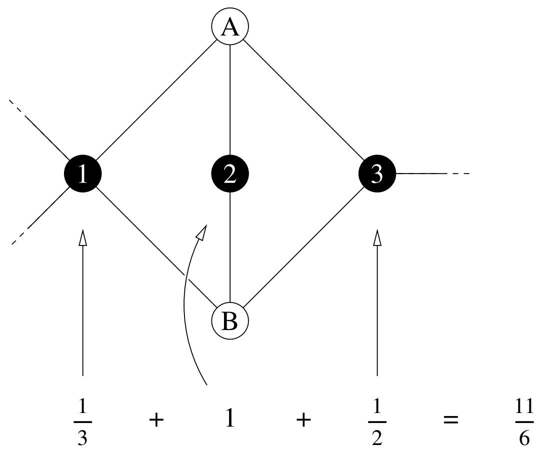


Figure 5: Authors A and B have coauthored three papers together, labeled 1, 2, and 3, which had respectively four, two, and three authors. The tie between A and B accordingly accrues weight $\frac{1}{3}$, 1, and $\frac{1}{2}$ from the three papers, for a total weight of $\frac{11}{6}$.

one of the present author's most frequent collaborators—with line thickness used to indicate the strength of the connections. Clearly there is considerable variation in connection strength. Over the entire physics network connection strengths range from a maximum of 34.0 to a minimum of about 0.01.

# 3 Additional results

The network of collaborations of mathematicians compiled by Grossman and Ion [9] and studied here has also been analyzed extensively by Grossman and collaborators [9, 10, 26] and by others [27]. Grossman [10] gives a number of results about the time evolution of the collaboration network. He notes that the rate of publication has increased slightly over the last fifty years or so, but there has been a much more striking increase in the level of collaboration. From the start of the period covered by the *Mathematical Reviews* data in 1940 until the end of the 1950s, less than half of all publishing mathematicians had coauthored a paper with another writer ever. Nowadays virtually all of them have. Presumably this trend reflects some combination of changes in social organization of the mathematics community, better communications, and possibly changes in the types of problems studied and approaches used, making modern mathematics more amenable to collaborative investigation.

The time evolution of coauthorship networks has also been investigated by the present author [15] and by others [11] in the context of tests of the "preferential attachment" hypothesis. Price [31] and later Barabási and Albert [32] have suggested that networks grow by the addition of connections in such a way

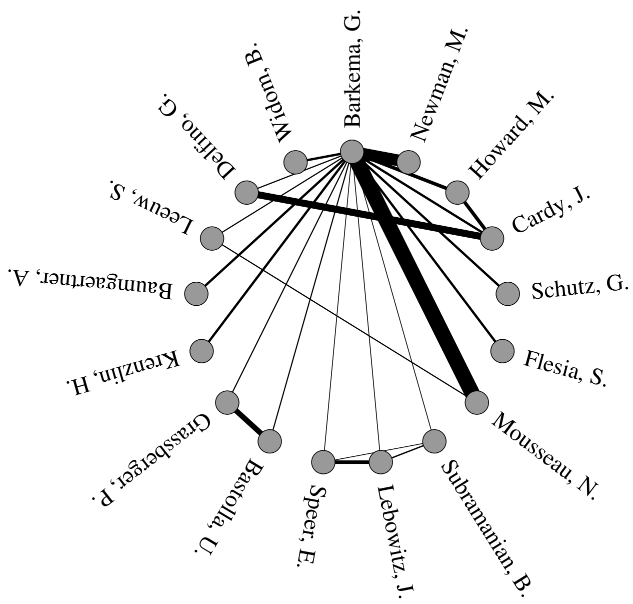


Figure 6: Gerard Barkema and his collaborators, with lines representing collaborations whose thickness is proportional to the estimate of collaboration strength defined defined in the text and illustrated in Fig. 5.

that the probability of an individual gaining a new connection is proportional to the number they already have. Since coauthorship networks are well time-resolved, as discussed in the preceding section, one can test this hypothesis by measuring the probability that a newly published paper contributes new connections to individuals as a function of the number of connections they already have. Refs. 11 and 15 tackle this measurement in slightly different ways, but both conclude that preferential attachment of an approximately linear variety is indeed taking place in collaboration networks.

A number of authors have also looked at "betweenness centrality" in coauthorship networks [14, 33–35]. The betweenness centrality of a node A in a network is defined to be the number of shortest paths between other pairs of nodes that pass through A [36]. It is regarded as a measure of the influence that individuals have over information flow between others. Individuals who act as brokers for information flow between their colleagues will have high betweenness scores. In 14 the present author found that betweenness scores vary widely from one individual to another in coauthorship networks, with a few having much higher scores than the majority. Later work by Goh *et al.* [34] shows that in fact the distribution of betweenness scores approximately follows a power law.

This appears to indicate that collaboration networks contain a small number of influential individuals, and many peripheral actors, a conclusion that is bolstered by the findings of Holme *et al.* [33], who show that collaboration networks are highly susceptible to the removal of the individuals with highest betweenness scores. One need only remove a few such individuals from the network, it turns out, to break the connection between others and fracture the network into disconnected parts. In a recent paper, Goh *et al.* [35] have extended their investigation of betweenness to the correlation between the betweenness scores of collaborators. They find that there is very little such correlation, implying that influential scientists are not collaborating preferentially with other influential scientists to any significant extent; the probability of your collaborator having a high betweenness appears to be not significantly greater if you yourself have a high betweenness than if you do not.

## 4 Conclusions

In this paper we have discussed the structure of three networks of scientific collaborations, as deduced from the pattern of coauthorships of published papers. The networks cover biomedical research, physics, and mathematics respectively, and the results reveal both similarities and differences between the different fields. All fields appear to have a broad distribution of the number of coauthors that an individual has, with most individuals having only a few coauthors, while a few have very many—hundreds or even thousands in some cases. Biological scientists tend to have significantly more coauthors than mathematicians or physicists, a result that reflects the labor-intensive, predominantly experimental direction of current biology. Other differences are less easily explained however. In biology, for instance, it is far less likely than in mathematics that two of your coauthors will also be coauthors of one another, a result that has yet to receive a clear explanation.

Coauthorship networks provide a copious and meticulously documented record of the social and professional networks of scientists. The results reported here represent only a small portion of what could be done with these data. Possible future directions for study might include tests for community structure or "invisible colleges" within the networks [3, 37] or further investigations of changes in collaboration patterns over time [10, 15], as well as other measurements not yet thought of. Coauthorship data represent a superb resource for the pursuit of questions such as these, and this author looks forward to future developments with interest.

## Acknowledgments

The author would particularly like to thank Paul Ginsparg for his help in obtaining the data used for this study. The data were generously made available by Oleg Khovayko, David Lipman, and Grigoriy Starchenko (Medline), Paul

Ginsparg and Geoffrey West (Physics E-print Archive), and Jerry Grossman *(Mathematical Reviews)*. The author would also like to thank Steve Strogatz for suggesting the "funneling effect" calculation of Section 2 and László Barabási, Paul Ginsparg, Jon Kleinberg, Sid Redner, Steve Strogatz, and Duncan Watts for useful comments and suggestions. This work was funded in part by a grant from Intel Corporation to the Santa Fe Institute's Network Dynamics Program, by the James S. McDonnell Foundation, and by the US National Science Foundation under grants DMS–0109086 and DMS–0234188.

## References

[1] D. J. de S. Price, Networks of scientific papers. *Science* **149**, 510–515 (1965).

[2] L. Egghe and R. Rousseau, *Introduction to Informetrics*. Elsevier, Amsterdam (1990).

[3] M. Girvan and M. E. J. Newman, Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99**, 8271–8276 (2002).

[4] H. Kretschmer, Coauthorship networks of invisible college and institutionalized communities. *Scientometrics* **30**, 363–369 (1994).

[5] O. Persson and M. Beckmann, Locating the network of interacting authors in scientific specialties. *Scientometrics* **33**, 351–366 (1995).

[6] G. Melin and O. Persson, Studying research collaboration using coauthorships. *Scientometrics* **36**, 363–377 (1996).

[7] Y. Ding, S. Foo, and G. Chowdhury, A bibliometric analysis of collaboration in the field of information retrieval. *Intl. Inform. and Libr. Rev.* **30**, 367–376 (1999).

[8] M. Bordens and I. Gómez, Collaboration networks in science. In H. B. Atkins and B. Cronin (eds.), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, Information Today, Medford, NJ (2000).

[9] J. W. Grossman and P. D. F. Ion, On a portion of the well-known collaboration graph. *Congressus Numerantium* **108**, 129–131 (1995).

[10] J. W. Grossman, The evolution of the mathematical research collaboration graph. *Congressus Numerantium* **158**, 202–212 (2002).

[11] A.-L. Barabási, H. Jeong, E. Ravasz, Z. Néda, A. Schuberts, and T. Vicsek, Evolution of the social network of scientific collaborations. *Physica A* **311**, 590–614 (2002).

[12] M. E. J. Newman, The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98**, 404–409 (2001).

[13] M. E. J. Newman, Scientific collaboration networks: I. Network construction and fundamental results. *Phys. Rev. E* **64**, 016131 (2001).

[14] M. E. J. Newman, Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E* **64**, 016132 (2001).

[15] M. E. J. Newman, Clustering and preferential attachment in growing networks. *Phys. Rev. E* **64**, 025102 (2001).

[16] M. E. J. Newman, Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).

[17] A. J. Lotka, The frequency distribution of scientific production. *J. Wash. Acad. Sci.* **16**, 317–323 (1926).

[18] W. Shockley, On the statistics of individual variations of productivity in research laboratories. *Proc. IRE* **45**, 279–290 (1957).

[19] H. Voos, Lotka and information science. *Journal of the American Society for Information Science* (July-August 1974), 270–272 (1974).

[20] M. L. Pao, An empirical examination of Lotka's law. *Journal of the American Society for Information Science* (January 1986), 26–33 (1986).

[21] T. Fenner, M. Levene, and G. Loizou, A stochastic evolutionary model exhibiting power-law behaviour with an exponential cutoff. Preprint cond-mat/0209463 (2002).

[22] I. de S. Pool and M. Kochen, Contacts and influence. *Social Networks* **1**, 1–48 (1978).

[23] S. Milgram, The small world problem. *Psychology Today* **2**, 60–67 (1967).

[24] J. Travers and S. Milgram, An experimental study of the small world problem. *Sociometry* **32**, 425–443 (1969).

[25] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Upper Saddle River, New Jersey (1993).

[26] R. de Castro and J. W. Grossman, Famous trails to Paul Erdős. *Mathematical Intelligencer* **21**, 51–63 (1999).

[27] V. Batagelj and A. Mrvar, Some analyses of Erdős collaboration graph. *Social Networks* **22**, 173–186 (2000).

[28] H. Kautz, B. Selman, and M. Shah, ReferralWeb: Combining social networks and collaborative filtering. *Comm. ACM* **40**, 63–65 (1997).

[29] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).

[30] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118 (2001).

[31] D. J. de S. Price, A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.* **27**, 292–306 (1976).

[32] A.-L. Barabási and R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

[33] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, Attack vulnerability of complex networks. *Phys. Rev. E* **65**, 056109 (2002).

[34] K.-I. Goh, E. Oh, H. Jeong, B. Kahng, and D. Kim, Classification of scale-free networks. *Proc. Natl. Acad. Sci. USA* **99**, 12583–12588 (2002).

[35] K.-I. Goh, E. Oh, B. Kahng, and D. Kim, Betweenness centrality correlation in social networks. *Phys. Rev. E* **67**, 017101 (2003).

[36] L. Freeman, A set of measures of centrality based upon betweenness. *Sociometry* **40**, 35–41 (1977).

[37] D. Crane, *Invisible colleges: Diffusion of knowledge in scientific communities.* University of Chicago Press, Chicago (1972).