The Kleiman graph

R. Alberich, J. M. Miret, J. Miro-Julia, F. Rosselló, S. Xambó

1 Introduction

A social network is a collection of people that may be related to each other through some specific relationship, like for instance friendship, collaboration in some kind of work, joint participation in some kind of event, etc. Such a network is usually represented by a graph, whose nodes and links denote, respectively, people and the existence of the relationship. Social networks have been much studied in social sciences, and some of its results have even left the academic realm and entered the popular culture through notions like the *six degrees of separation* [11], which asserts that any two randomly selected people in the world can be "connected" through a chain of a few (namely, six) intermediate acquaintances.

A popular type of social networks are the *collaboration networks*, whose nodes and links represent, respectively, people and some kind of collaboration between them. A well-known example of collaboration network is the *Movie Actors network*, also dubbed the *Hollywood network*. In it, nodes represent actors and actresses, and a link is added between two nodes when they have jointly appeared in the same film. All information concerning this network is accessible at the *Internet Movie Database* [12], and it has been studied from a mathematical point of view [1, 21, 22]. This network is the basis of the popular *Kevin Bacon game* [20], which consists of trying to connect any given actor or actress to Kevin Bacon through the shortest possible path of collaborations in films.

Closer to us, *scientific collaboration networks* have also been thoroughly studied in the last years. In such a network, nodes represent scientists and links denote the coauthorship of a scientific piece of work contained in some database. Among us, mathematicians, the most famous collaboration network is the *Erdös collaboration graph* [7], the mathematicians' collaboration network around the prolific Hungarian mathematician Paul Erdös, dead in 1996, who published over 1500 papers with 492 coauthors, more than any other mathematician in history. This network has been studied both from the formal [5, 10] and the gossipy [6] points of view, and it is used to compute *Erdös numbers*: Erdös himself has Erdös number 0, his collaborators have Erdös number 1, the coauthors of the latter that have not collaborated with Erdös have Erdös number 2, and so on.

The study of the whole mathematical research collaboration graph, based on the more than 1.6 million entries in the *Mathematical Reviews* database, has been recently initiated by J. Grossman [9] as a spin-off of his study of the Erdös collaboration graph. Also, M. E. J. Newman has studied in detail the scientific collaboration networks corresponding to several databases, namely MEDLINE (biomedical research papers in refereed journals), SPIRES (preprints and published papers in high-energy physics), NCSTRL (preprints in computer science), and Los Alamos e-Print Archive (preprints in physics) [15–19], while A. L. Barabási *et al* have studied the networks based on two databases containing articles on mathematics and neuro-science respectively, published in all relevant journals in these fields in the period 1991-98 [3].

All collaboration networks studied so far, as well as most social networks, present the same basic features: (a) on average, every pair of nodes can be connected through a short path within the network; (b) the probability that two nodes are linked is much greater if they share a neighbor; and (c) the fraction of nodes with k neighbors decays roughly as a function of the form $k^{-\tau}$ for some positive exponent τ. A network satisfying properties (a) and (b) is called a *small world* [21, 22], and if it satisfies (c) then it is called *scale-free* [1, 2].

For instance, in the study of J. Grossman [9] on the mathematical research collaboration network based on the whole set of entries of *Mathematical Reviews* up to May 2000, the following figures were obtained:

– On average, each pair of mathematicians can be connected through a path of less than 7 intermediate coauthors.
– The probability that two authors sharing a collaborator (possibly in different papers) have collaborated in some paper is 0.15, while the probability that two mathematicians picked at random have collaborated in some paper is 0.0000087. This last figure is consequence of the fact that, although this network covers 337 454 mathematicians, on average each one has only 2.94 collaborators.
– The fraction of mathematicians with k collaborators decays roughly as $k^{-2.81}$.

The figures obtained by Barabási *et al* are slightly different, and they will be recalled elsewhere in this report.

On occasion of Steven Kleiman's 60th birthday, we considered a good tribute to him to make a statistical analysis of the features of the collaboration network in the field of enumerative geometry. This analysis has two purposes: on the one hand, to support everybody's perception that Kleiman is the most relevant member of the enumerative geometry world, and, on the other hand, to extract some conclusions about the collaboration behavior of the enumerative geometry community when compared with other scientific communities, and specially with the mathematicians' community at large. This second aspect of our work is a contribution to the problem posed by J. Grossman [9], who asked for a comparison of the properties of the mathematical research collaboration graph with those of weak subgraphs of it restricted to some specific subjects or branches of mathematics.

We have decided to carry out our study of the enumerative geometry collaboration network at two stages: one covering only until 1992, and the other corresponding to its whole history until today. The reason is that we believe that with the discovery around 1992 of the connection between enumerative geometry and particle physics, a new bunch of scientists, mainly differential geometers and physicists, entered this field, which, until that moment, had been mainly dominated by algebraic geometers. This study seems thus a good occasion to check wether this irruption of mathematicians from other areas affected the structure of the field.

This work has been partly supported by the DGES, grant BFM2000-1113-C02-01.

2 The Enumerative Geometry world

To build up the enumerative geometry network, we have used all entries filed in the *Mathematical Reviews* database (available on the World Wide Web as *MathSciNet* [14]) under the AMS codes 14N10 (Enumerative Geometry) and 14C17 (Intersection Theory). From these entries, we have derived a collaboration network with nodes the authors of the corresponding books and articles (*papers*, for short) and links representing coauthorship. We have actually built up *two* such collaboration networks: one, which we will refer to as *EG-92*, corresponding only to entries dated 1992 or before, and the other, *EG-02*, defined by the whole set of enumerative geometry and intersection theory entries in the database.

We want to make clear that we have not considered all papers written by the authors involved in this study, but only those that have been explicitly classified as belonging to the fields of enumerative geometry or intersection theory by their authors or their reviewers. In a similar way, we have not taken into account all enumerative geometry papers written so far, but only those collected in the *Mathematical Reviews* database.

We should devote a little time to discuss what an "author" means in this context. From the database it is straightforward to extract the *names* that have signed papers in it, but there is not a one-to-one relation between names and authors. For instance, one of the authors of this report, F. Rosselló, appears in different entries as F. Rosselló, Francesc Rosselló, Francesc A. Rosselló and F. Rosselló-Llompart: four names for a single author. Even Kleiman does not escape this effect: he has signed papers as S. Kleiman, Steven Kleiman and Steven L. Kleiman. The opposite effect happened when, for instance, Susan Jane Morris married William C. Colley III and became Susan Jane Colley; fortunately, she did not sign any enumerative geometry paper under her maiden name.

This lack of one-to-one correspondence between names and authors is usually avoided by systematically identifying two names as the same author when they have the same surname and the same first initial, or the same surname and the same two initials. In our study, and since the number of entries under consideration allowed it, we took the safer way of solving case by case all indeterminacies using our personal knowledge of the mathematicians involved in it as well as the tools available for that purpose on *MathSciNet*, and we are thus quite sure that we have always identified two names referring to the same person and we have never confused two people for one.

3 Analysis of the network

From the data on intersection theory and enumerative geometry papers extracted from the *Mathematical Reviews* database, we have built up two bipartite graphs, one from those entries up to 1992 and the other one with all entries, with nodes corresponding to either authors or papers and edges from every author to all papers he or she has written. We have extracted then from these bipartite graphs the EG-92 and EG-02 networks, as their respective projections on their sets of nodes, and we have used PAJEK, a program for large network analysis [4], and our own programs to compute the relevant statistical values, which are numerically summarized in Tables 1 and 3 and discussed in the next subsections. To ease the task of comparing the results obtained for the enumerative geometry world with those obtained in other analysis of other mathematicians' collaboration networks, we include in these tables two extra columns: one under the heading of *whole MR*, with the values obtained by J. Grossman in his study [9, 8] of the mathematical research collaboration graph extracted from the whole set of entries in the *Mathematical Reviews* database, and another under the heading *Barabási*, containing the values obtained by A. L. Barabási *et al* in their study [3] based on all papers published in all major mathematical journals between 1991 and 1998, covering 70 901 papers by 70 975 authors. This last work does not include data on the corresponding bipartite graph, and therefore this column is only included in Table 3.

	1992	2002	whole MR
Number of authors:	273	545	337 454
Number of papers:	441	891	$\sim 1\,600\,000$
Mean papers per author:	2.14	2.24	6.87
Authors with only one paper:	66%	61.6%	42%
Mean authors per paper:	1.32	1.37	1.45
Papers with a single author:	72.3%	65.8%	65.8%
Distribution of papers per author:	$P_a(k) \sim k^{-1.65}$	$P_a(k) \sim k^{-1.86}$	
Distribution of authors per paper:	$P_p(k) \sim 10^{-0.63k}$	$P_p(k) \sim 10^{-0.60k}$	

Table 1. Basic data on authors and papers.

3.1 The bipartite graphs

The bipartite graph summarizing the EG-92 network contains 273 nodes corresponding to authors and 441 nodes corresponding to papers, and 585 edges going from authors to their papers, while the bipartite graph summarizing the EG-02 network contains 545 nodes corresponding to authors and 891 nodes corresponding to papers, and 1225 edges going from authors to their papers.

In 1992, an enumerative geometer had typically published 2.14 papers in this field, while today he/she has typically published 2.24 papers. Actually, up to 1992, 80% of the enumerative geometers had published only one or two papers in this field, and this percentage is today 77%. But, contrary to what these two pieces of data seem to hint, the productivity rate in the field of enumerative geometry has decreased from 1992 to 2002. To see this, let us consider the distribution $P_a(k)$ of papers per author in these bipartite graphs; i.e., let $P_a(k)$ denote the probability that an author has published k papers. Performing a linear regression of $\log(P_a(k))$ on $\log(k)$, we obtain that both in EG-92 and in EG-02 these distributions follow a power-law, namely

$$P_{a,1992}(k) \sim 0.35 \times k^{-1.65}, \qquad P_{a,2002}(k) \sim 0.51 \times k^{-1.86}.$$

These distributions are jointly represented in Fig. 1. Notice the steeper slope of the approximation to the distribution of papers per author in 2002, which represents a lower probability of high productivity. This shows that newcomers compensate (negatively) for the fact that senior authors in the field have increased their number of papers. Of course, there are exceptions to this trend both among the seniors and the newcomers: there are prolific authors in the period up to 1992 that have published only one enumerative geometry paper in the last decade, and there are newcomers in the field, like S.-T. Yau (well, we resist to consider him a newcomer), B.-H. Lian and G. Tian, that had not published any paper in the field before 1992 but they have published 10 or more enumerative geometry papers in the period 1993-2002.

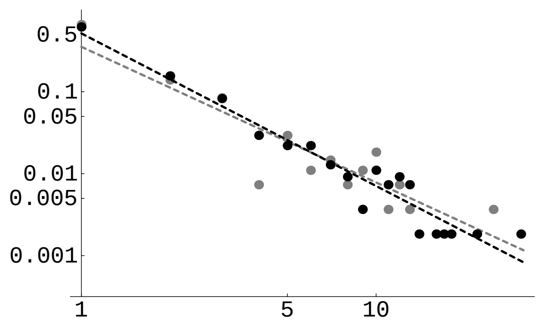

Fig. 1. Distribution of papers per author in the enumerative geometry bipartite graphs. The horizontal axis corresponds to the number of papers, while the vertical axis represents the frequency of authors that have published those many books. Note that the scales on both axis are logarithmic. The dashed lines represent the log-log regressions, and the dots represent the actual frequencies; the grey lines and dots correspond to 1992, while the black ones correspond to 2002.

S. Kleiman is in both periods under study the mathematician that has published more enumerative geometry papers: 25 until 1992 and 6 more in the last decade. Table 2 shows the ten most prolific authors in the field up to 1992 and today. As far as the members of the list corresponding to 2002 in this table, it is interesting to point out that, as we have already mentioned, S.-T. Yau had not published any paper on enumerative geometry before 1992, while P. Aluffi, who had recently finished his PhD thesis under Fulton, had only published 6.

1992		2002	
	Number of papers		Number of papers
S. Kleiman	25	S. Kleiman	31
W. Vogel	13	W. Vogel	22
P. Le Barz	12	W. Fulton	18
R. Speiser	12	P. Aluffi	17
S. Xambó	11	I. Vainsecher	16
I. Vainsecher	10	P. Le Barz	14
H. Gillet	10	R. Speiser	13
R. Piene	10	S.-T. Yau	13
W. Fulton	10	R. Piene	13
D. Laksov	10	H. Flenner	13

Table 2. Most prolific enumerative geometry authors.

Up to 1992, a typical enumerative geometry paper had 1.32 authors, and today it has 1.37 authors. Actually, in 1992, from the 441 considered papers, 319 were single-authored, which yields a 72.3% and only one had more than 3 authors: namely, "Rational points" by G. Faltings, G. Wüstholz, F. Grunewald, N. Schappacher, and U. Stuhler (which, by the way, is a collection of papers from a seminar held at the Max-Planck Institut in 1983/84, published by Friedr. Vieweg & Sohn in 1992, and it has 14C17 only as secondary code). Today, 586 from the 891 papers are single-authored, a 65.8% (exactly the same as in the whole mathematical research collaboration network!), and there are 14 with more than 3 authors, including two more papers with 5 authors. But let us mention that, from these 13 new papers with many coauthors, both papers with 5 authors and 7 papers with 4 authors are published in physics journals or books, and the remaining 4 have the codes 14N10 or 14C17 as secondary.

Notice that the figures reported in the previous paragraph imply that only a 59.3% of the enumerative geometry papers published during the last decade are single-authored, and they seem to hint that there has been a tendency to increase the number of authors per paper from 1992 to 2002. If we look at the distribution $P_p(k)$ of authors per paper in the bipartite graphs under consideration, i.e., if $P_p(k)$ denotes the probability that a paper has k authors, and if we perform as before a linear regression of $\log(P_p(k))$ on $\log(k)$, we obtain that these distributions follow exponential laws, namely

$$P_{p,1992}(k) \sim 3.58 \times 10^{-0.64k}, \qquad P_{p,2002}(k) \sim 3.22 \times 10^{-0.60k};$$

they are jointly represented in Fig. 2. From these distributions we can conclude indeed that the tendency during the last decade in the enumerative geometry world has been towards increasing the number of authors per paper.

Thus, enumerative geometers do not seem prone to collaborate in large groups. This tendency is shared with the rest of mathematical fields: only around a 1.7% of mathematical papers have more than 3 authors [8, 13], while . But although the average number of authors per enumerative geometry paper has increased along the last decade, this increase is mainly due to the arrival to the field in the last decade of physics-related papers on mirror symmetries and the like, and even with their contribution the percentage of papers with more than three authors is slightly smaller than that of the whole mathematicians' collaboration network.

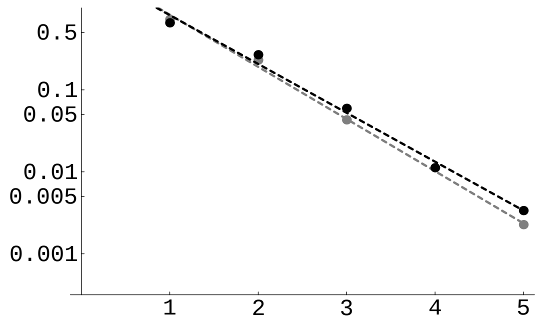

Fig. 2. Distribution of authors per paper in the enumerative geometry bipartite graphs. The horizontal axis corresponds to the number of authors, while the vertical axis represents the frequency of papers with those many authors. Note that the scale on the vertical axis is logarithmic, while the scale on the horizontal axis is linear. The dashed lines represent the regressions, and the dots represent the actual frequencies; the grey lines and dots correspond to 1992, and the black ones correspond to 2002.

3.2 Basic data

Our EG-92 network has $N_{92} = 273$ nodes (authors) and $M_{92} = 130$ links, i.e. pairs of authors that in 1992 had jointly written some paper on enumerative geometry. On the other hand, the EG-02 network has $N_{02} = 545$ nodes and $M_{02} = 326$ links.

	1992	2002	whole MR	Barabási
Authors:	273	545	337 454	70 975
Mean collaborators per author:	0.95	1.19	2.94	3.9
Size of core component:	15.8%	17.6%	61.7%	71%
Maximum distance in the core:	8	11	27	
Mean distance in the core:	3.9	5	~ 7.5	9.5
Clustering coefficient:	0.37	0.38	0.15	0.59
Distribution of coauthors:	$P(k) \sim k^{-1.36}$	$P(k) \sim k^{-1.55}$	$P(k) \sim k^{-2.81}$	$P(k) \sim k^{-2.4}$

Table 3. Basic data on the enumerative geometry collaboration network.

We would like to mention that the actual number of collaborations in EG-92 is 169 and in EG-02, 488, but these values count *all* collaborations in the enumerative geometry history, and while, for instance, in EG-02 there are 243 pairs of authors that have only collaborated in one paper, which amounts roughly to the 75% of all pairs of coauthors, other pairs have collaborated quite often: for instance, B.-H. Lian and S.-T. Yau have jointly written 10 enumerative geometry papers (all of them after 1992), and H. Gillet and C. Soulé have written 8 joint papers, 7 of them before 1992. In this sense, Kleiman is not prone to repeat collaborations: its greatest number of collaborations with another author in this field is three (with R. Speiser).

The number of authors that have collaborated with a given author in some paper is given by the *degree* of this author in the network. The average value for this degree can be computed as twice the number of links in the network divided by its number of nodes. In our EG collaboration networks, these average degrees are

$$\frac{2M_{92}}{N_{92}} = 0.95, \qquad \frac{2M_{02}}{N_{02}} = 1.19.$$

This shows that an enumerative geometer had collaborated on average with one coauthor before 1992, and today this average has increased to around 1.2. You should compare these figures with the average number of coauthors in mathematics at large reported by Grossman and by Barábasi *et al*, which are 2.94 and 3.9, respectively.

As a matter of fact, let us point out that in GE-92 there are 135 authors (around 49.5% of the whole set of authors) that had written all their enumerative papers alone, and this figure reaches today 221 authors (40.5% of the authors). In the whole area of mathematics, according to J. Grossman, only (only?) 25% of mathematicians have not collaborated with anybody else. All this shows again that enumerative geometry is not a research field prone to collaboration.

Anyway, both up to 1992 and today, the author in the field with the greatest number of collaborators is again S. Kleiman, with 11 collaborators until 1992 and 13 coauthors until 2002. Table 4 shows the eight authors with the greatest number of collaborators until 1992 and 2002 (after the authors listed in 1992, there is a group of nine authors with three collaborators each).

1992		2002	
	Number of coauthors		Number of coauthors
S. Kleiman	11	S. Kleiman	13
S. Xambó	6	W. Vogel	12
A. Lascoux	5	W. Fulton	8
D. Laksov	5	R. Achiles	6
R. Piene	4	C. de Concini	6
E. Ballico	4	H. Flenner	6
W. Fulton	4	I. Sols	6
		S. Xambó	6

Table 4. Authors with the highest numbers of collaborators.

We shall discuss the distribution of collaborators in EG-92 and EG-02 in §3.6.

3.3 The core component

Two nodes in a network are said to be *connected* when there is at least one path in the network, made of consecutive links, that connects them. In a collaboration network, this means that two nodes are connected when they can be linked through a path of intermediate collaborators. As mentioned above, finding such a path, and more specifically the shortest one, between any movie actor or actress and Kevin Bacon in the Hollywood network, is the goal of the Kevin Bacon game.

Actually, two nodes in a collaboration network need not be connected, and in particular there need not exist a path of copartnerships connecting every actor in the world to Kevin Bacon. But, in all large enough real-life networks, a large fraction of nodes are connected to each other. More specifically, large collaboration networks (and other large social networks) usually contain a very large subset of nodes —from 60% to 90% of all nodes— that are connected to each other: when this happens, this large subset of nodes with their corresponding links is called the *giant* or *core component* of the network. For instance, the largest connected component in the mathematical research collaboration network has 208 200 nodes, which corresponds to a 61.7% of the set of authors. As its name hints, this core component is considered to be the core of the network, and nodes outside it are understood as not being part of the "gist" of the network.

Both enumerative geometry networks under consideration contain large connected components, but they are by no means giant. The largest connected component in EG-92 has 43 authors, spanning only a 15.75% of the total set of authors, while in EG-02 it has 96 authors, which amounts to a 17.61%. Thus, in the enumerative geometry world the core components are much smaller than in other scientific collaboration networks. A reason for this fact is the existence of large sets of isolated authors, covering about half the network in 1992 and two fifths today: in the mathematical research collaboration network there are around 84 000 mathematicians that have not written any joint paper, which amounts to only a 25% of the mathematical community. Another reason is that the second largest connected components in these networks have 12 and 13 authors, respectively, and therefore they are quite larger than in other scientific collaboration networks, where they cover between a 0.002% and a 0.2% of the whole set of authors: in the mathematical research collaboration network, the second largest connected component has 39 nodes, a 0.011% of the mathematical community

By the way, if somebody is interested, the second largest connected component in EG-92 has merged into the core component of EG-02. Perhaps it's time for Sheldon Katz, David Cox or some other member of this second largest component in EG-02 to write a paper with somebody in the core component!

The members of the core components of the enumerative geometry world in 1992 and in 2002 are listed in Table 6 at the end of this paper, and the corresponding graphs are represented in Fig. 4 and 5 also at the end.

3.4 Separation

The *distance* between two connected nodes in a network is defined as the length (the number of links) of the shortest path connecting them, i.e., the least number of links we have to traverse in order to move from one node to the other within the network. Notice that the number of links in a path is equal to the number of intermediate nodes plus one, and thus in a collaboration network we could also say that the distance between two connected nodes is the least number of intermediate collaborators visited by a path connecting them, plus one. For instance, the Kevin Bacon game asks for the distance of any actor or actress to Kevin Bacon in the Hollywood network, as the least number of intermediate co-partners plus one linking that actor or actress to Kevin Bacon, and the Erdös number of a mathematician is simply his/her distance to P. Erdös in the mathematicians' collaboration network. The graph shown in Fig. 5 will allow you to play the *Steven Kleiman game* and to compute *Kleiman's numbers*, although I am afraid that this amusement will probably never become as popular as the Kevin Bacon game or the Erdös numbers game.

Since distances only make sense within connected components (or, if you want, two nodes in different connected components are at an infinite distance), and in small connected components distances are, of course, consistently small, we have only calculated the distances between all pairs of nodes in the core components of EG-92 and EG-02.

In 1992, the greatest distance between two nodes in the core of the enumerative geometry world, its *diameter* in the usual network-theoretical terminology, was 8, while in 2002 the corresponding diameter was 11: the latter corresponds to the distance from V. V. Batyrev to J. Michel. That means that in 1992 there always was a chain of at most 7 collaborators connecting any two authors in the core component, while in 2002 this figure had increased to 10.

We have also computed the mean of all distances in the core components, which provides the average separation of two authors in them. The value of this average separation in 1992 was 3.9, while today it is 5. Thus, on average, any pair of enumerative geometry authors in the core component can be connected today through a path of at most four consecutive partners, while in 1992 it could be done on average with at most three consecutive coauthors.

In both periods of time, the *center* of the core component, the author that minimizes the sum of the distances from it to all other nodes in the component, turns out to be S. Kleiman. Table 5 shows the ten most centered members of the core components in 1992 and in 2002, with their average distances to other authors. The center of today's second largest connected component is Sheldon Katz.

1992		1992	
	Average distance		Average distance
S. Kleiman	2.19	S. Kleiman	2.91
D. Laksov	2.76	W. Fulton	3.20
S. Xambó	2.78	B. Ulrich	3.42
R. Piene	2.83	R. Speiser	3.52
R. Speiser	2.88	S. Xambó	3.58
W. Fulton	2.93	D. Laksov	3.61
S. A. Strømme	2.92	S. A. Strømme	3.61
A. Hefez	2.95	R. MacPherson	3.64
R. MacPherson	3.00	J. Lipman	3.66
J. Lipman	3.14	R. Piene	3.68

Table 5. Most centered authors in the core component.

3.5 Clustering

In most social networks, two nodes that are linked to a third one have a higher probability to be linked to each other: for instance, two acquaintances of a given person probably know each other. This effect is measured using the *clustering coefficient* of the network, that is defined as follows. Let an *angle* be an ordered 3-tuple of pairwise different nodes (u, v, w) such that u and w are neighbors of v, and let a *triangle* be a set of 3 nodes $\{u, v, w\}$ where each node is neighbor of the other two. In other words, an angle corresponds to a node connected to two other nodes, and a triangle corresponds to three nodes connected to each other. We can now define the clustering coefficient C as

$$C = \frac{3 \times \text{number of triangles in the network}}{\text{number of angles in the network}}.$$

Note that each triangle contributes three angles, so in the formula above the number of triangles must be multiplied by 3. This value C represents the probability that two neighbors of an arbitrary node are linked.

All collaboration networks considered so far, and in general most social networks, have large clustering coefficients, which tells us that a large fraction of collaborators of a scientist collaborate with each other. This large clustering, together with a low value of the average distance between connected nodes, is taken as the definition of *small-world* networks [21]. Actually, the word "large" means large compared to the probability that two nodes picked at random in the network are linked to each other: in a network of N nodes and M links, this probability is

$$C_{\text{random}} = \frac{2M}{N(N-1)}.$$

Measured values in collaboration networks are usually "large" in the sense that they are several orders of magnitude larger than the corresponding predicted value of C_{random}, which means that the probability that two authors have jointly written some paper if they share a collaborator, is several orders of magnitude greater than if both authors are picked at random. For instance, the clustering coefficient of the MEDLINE network is 0.066, which seems small, but it becomes very large when compared with the value 0.0000042 of the probability that two randomly chosen authors in it have collaborated. In a similar way, the clustering coefficient of the mathematical research collaboration network is 0.15,[1] but the probability that two randomly chosen mathematicians have collaborated is 0.0000087, more than 17 000 times lower. You might draw some consequences from the fact that the clustering coefficient of the mathematics collaboration network is similar to that of the Hollywwod graph, which is 0.199.

The enumerative geometry collaboration network is highly clustered, but not too much: its clustering coefficient in 1992 was 0.37, while today it is 0.38. This values are, indeed, higher than the 0.15 figure quoted above, but they are not much greater than the probability that two authors in the network picked at random are linked to each other, which is 0.058 in 1992 and 0.029 in 2002.

In other words, in 1992, the probability that two mathematicians had jointly authored an enumerative geometry paper was five times greater if they shared a coauthor in this field, and in 2002 this probability was more than twelve times greater in the presence of a collaborator in common, while in mathematics in general, this factor is more than 17 000. This shows that people working in enumerative geometry do not tend to form working groups with a high degree of collaboration in them, and that their collaborations do not spread among their coauthors. This adds to the small average number of collaborators and the small average number of authors of papers to make us conclude that enumerative geometry is quite an individualistic field.

One can also compute the clustering coefficient of a given node v_0 in a network as

$$C_{v_0} = \frac{2 \times \text{number of triangles in the network containing } v_0}{\text{number of angles in the network of the form } (u, v_0, w)}.$$

In this case, the numerator is multiplied by two because every triangle $\{u, w, v_0\}$ containing v_0 only contributes the angles (u, v_0, w) and (w, v_0, u) to the denominator. This coefficient C_{v_0} measures the fraction of neighbors of node v_0 that are linked. If k_{v_0} denotes the degree of v_0 and N_{v_0} denotes the number of links between the neighbors of v_0, then we also have that

$$C_{v_0} = \frac{2N_{v_0}}{k_{v_0}(k_{v_0} - 1)}.$$

If $k_{v_0} \leq 1$, then C_{v_0} is taken to be 0.

A high clustering coefficient for an author in a scientific collaboration network means a high tendency of its coauthors to collaborate among themselves, but this value must be taken with care. For instance, a clustering coefficient value of 1 can simply mean that this author has only contributed one paper to the network: it is clear then that his collaborators have also collaborated with each other.

Kleiman's clustering coefficient in 1992 was 0.072 and in 2002 it is 0.064: thus, the tendency of his collaborators to collaborate among themselves is only double than if they decided to collaborate at random.

3.6 Distribution of the number of partners

An interesting statistical datum about a collaboration network is the distribution $P(k)$ of degrees in it. For every positive integer k, let $P(k)$ denote the fraction of nodes in a given network that

[1] As it is recalled in Table 3, Barabási *et al* report a clustering coefficient of 0.59 for their mathematicians' collaboration network, but this value cannot be compared with the ones given here, because they use a different formula to compute it: namely, as the mean of the clustering coefficients of all nodes (see below).

have degree k. In a random network with N nodes and M links, the expected value for $P(k)$ follows a binomial distribution

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}, \qquad p = 2M/N.$$

Against these values, it has been observed [1, 2] that in all collaboration networks considered so far the distribution $P(k)$ has a tail that follows either a power law

$$P(k) \sim k^{-\tau}$$

for some constant, positive exponent τ, or a power law form with an exponential cutoff

$$P(k) \sim k^{-\tau} 10^{-k/c}$$

where τ and c are two positive constants, and c is large. While the power law component of such a distribution allows the existence of a non-negligible number of nodes with high degree, the cutoff prevents the existence of nodes with very high degree. In most collaboration networks, this cutoff is explained because the collaboration under consideration can only take place in a finite amount of time (for instance, the professional lifetime of an actor or a scientist), which makes unplausible the existence of nodes with a number of collaborators greater than some reasonable upper bound.

For instance, the degree distribution in the whole mathematical research collaboration network follows a power-law $P(k) \sim k^{-2.81}$, while the degree distribution in the mathematics network studied by Barabási et al follows a power-law $P(k) \sim k^{-2.4}$.

When we tested how our data fit to a power-law distribution with an exponential cutoff, we found that the degree distributions in EG-92 and in EG-02 follow power-laws without cutoff, namely

$$P_{1992}(k) \sim k^{-1.36}, \qquad P_{2002}(k) \sim k^{-1.55};$$

these distributions are jointly represented in Fig. 3. The absence of cutoff is due both to the small absolute value of the exponent of the power-law and to the small size of the network. From the slopes of these distributions we conclude that the tendency during the last decade in the enumerative geometry world has been towards *decreasing* the average number of collaborators.

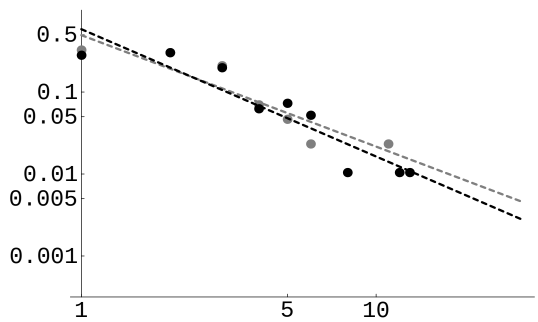

Fig. 3. Distribution of collaborators per author in the enumerative geometry collaboration graphs. The horizontal axis corresponds to the number of collaborators, while the vertical axis represents the frequency of authors with those many collaborators. Note that the scales on both the vertical and the horizontal axis are logarithmic. The dashed lines represent the log-log regressions, and the dots represent the actual frequencies; the grey lines and dots correspond to 1992, while the black ones correspond to 2002.

A value of τ smaller than 2 means that the average properties of the network are dominated by those of the few authors with a large number of collaborators. This increases the importance of the role of Kleiman in this network.

4 Conclusion

The analysis of the enumerative geometry collaboration network shows that Steven Kleiman is, both today and in 1992, the most prolific enumerative geometry author, the mathematician with the greatest number of coauthors in this field, and the center of the enumerative geometry world. Thus, he is without any doubt the most important mathematician in this field. But, on the other hand, he seems not to like to repeatedly collaborate with the same colleagues and their collaborators do not tend to work together. Nobody's perfect!

On the other hand, the overall picture of the statistics about authoring of papers in enumerative geometry shows an area with a low tendency to collaboration, even lower than in mathematics at large, and only the incorporation during the last decade of physicists, traditionally more prone to collaboration, has made its collaboration rate increase.

Of course, this lack of collaboration only refers to coauthorship of enumerative geometry papers. There are other measures that we have not taken into account, like meeting at conferences or joint participation at seminars, that could show a completely different picture. Moreover, enumerative geometers are quite a friendly bunch, and we are sure that many members of the enumerative geometry community under consideration have extensively collaborated with many other people in other fields (for instance, one of the authors of this report, F. Rosselló, has only one collaborator in enumerative geometry, S. Xambó, but he has jointly written papers with up to twenty coauthors in many other fields).

References

1. L.A.N. Amaral, A. Scala, M. Barthelemy, H.E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences U.S.A.* **97**, 21 (2000), pp. 11149-11152.
2. A.-L. Barábasi, R. Albert. Emergence of scaling in random networks. *Science* **286** (1999), pp. 509–512.
3. A.-L. Barábasi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek. Evolution of the social network of scientific colaborations. <http://www.arxiv.org/abs/cond-mat/0104162> (2001).
4. V. Batagelj, A. Mrvar. PAJEK—program for large network analysis. *Connections* **21** (1998), pp. 47–57. The program is available at <http://vlado.fmf.uni-lj.si/pub/networks/pajek>.
5. V. Batagelj, A. Mrvar. Some analyses of Erdös collaboration network. *Social Networks* **22**, 2 (2000), pp. 173–186.
6. R. De Castro, J.W. Grossman. Famous Trails to Paul Erdös. *The Mathematical Intelligencer* **21** (1999), pp. 51–63.
7. J.W. Grossman. "The Erdös Number Project" <http://www.oakland.edu/~grossman/erdoshp.html>. (Last visited, May 2002)
8. J.W. Grossman. "Facts about Erdös numbers and the Collaboration Graph." <http://www.oakland.edu/~grossman/trivia.html> (Last visited, May 2002).
9. J.W. Grossman. "The evolution of the mathematical research collaboration network." Preprint (2002). Available at <http://www.oakland.edu/~grossman/eddie.pdf>.
10. J.W. Grossman, P.D.F Ion. "On a portion of the well-known collaboration graph". Congresus Numerantium **108** (1995), 129–131.
11. J. Guare. *Six Degrees of Separation.* Vintage (New York, 1990).
12. "Internet Movie Database" <http://www.imdb.com> (Last visited, May 2002).
13. S. Liberman, K. B. Wolf. Bonding numbers in scientific disciplines. *Social Networks* **20** (1998), 239–246.

14. “MathSciNet, *Mathematical Reviews* on the Web.” <http://www.ams.org/mathscinet> (Last visited, May 2002).
15. M.E.J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences U.S.A.* **98**, 2 (2001), pp. 404–409.
16. M.E.J. Newman. Scientific collaboration networks I. Network construction and fundamental results. *Physical Review E* **64** (2001), 016131.
17. M.E.J. Newman. Scientific collaboration networks II. Shortest paths, weighted networks, and centrality. *Physical Review E* **64** (2001), 016132.
18. M.E.J. Newman. Clustering and preferential attachment in growing networks. *Physical Review E* **64** (2001), 025102.
19. M.E.J. Newman. Who is the best connected sicentist? A study of scientific coauthorship networks. <http://www.arxiv.org/abs/cond-mat/0011144 v2>.
20. B. Tjaden, G. Wasson. “The Oracle of Bacon at Virginia”. <http://www.cs.virginia.edu/oracle/>. (Last visited, May 2002).
21. D.J. Watts. *Small Worlds*. Princeton Univ. Press (1999).
22. D.J. Watts, S.H. Strogatz. Colective dynamics of “small-world” networks. *Nature* 393 (1998), pp. 440-442.

1992	2002	
Altman, A. B.	Achilles, R.	Lehn, M.
Arrondo, E.	Aliffi, D.	Lipman, J.
Ballico, E.	Altman, A. B.	MacPherson, R.
Bertin, J.	Arrondo, E.	Mallavibarrena, R.
Casas, E.	Ballico, E.	Manaresi, M.
Collino, A.	Barth, W.	McLaughlin, J. E.
De Concini, C.	Batyrev, V. V.	Michel, J.
Ellia, Ph.	Bauer, Th.	Miret, J. M.
Ellingsrud, G.	Belorousski, P.	Miyazaki, C.
Fulton, W.	Bertin, J.	Muñoz, V.
Ghione, F.	Bini, G.	Murre, J. P.
Gianni, P.	Bresinsky, H.	Nakamaye, M.
Goresky, M.	Buch, A. S.	O'Carroll, L.
Hefez, A.	Campana, F.	Pandharipande, R.
Holme, A.	Casas, E.	Parusiński, A.
Józefiak, T.	Ciocan-Fontanine, I.	Patil, D. P.
Kempf, G.	Collino, A.	Peters, C.
Kleiman, S.	Chadzyński, J.	Piene, R.
Laksov, D.	de Concini, C.	Polito, M.
Lascoux, A.	Demailly, J.	Pragacz, P.
Lazarsfeld, R.	Dutta, S. P.	Procesi, C.
Lipman, J.	Ein, L.	Pruschke, T.
MacPherson, R.	Ellia, Ph.	Ratajski, J.
Mallavibarrena, R.	Ellingsrud, G.	Roberts, J.
Miret, J. M.	Fantechi, B.	Rosselló, F.
Murre, J. P.	Flenner, H.	Sacchiero, G.
Parusiński, A.	Fulton, W.	Schützenberger, M.-P.
Piene, R.	Göttsche, L.	Simonis, J.
Pragacz, P.	Ghione, F.	Sols, I.
Procesi, C.	Gianni, P.	Speiser, R.
Roberts, J.	Goresky, M.	Stückrad, J.
Rosselló, F.	Graber, T.	Strømme, S. A.
Sacchiero, G.	Hefez, A.	Sturmfels, B.
Schützenberger, M.-P.	Hernández, R.	Tai, H.-S.
Sols, I.	Hochster, M.	Thorup, A.
Speiser, R.	Holme, A.	Traverso, C.
Strømme, S. A.	Huibregtse, M. E.	Trung, N. V.
Tai, H.-S.	Huneke, C.	Tworzewski, P.
Thorup, A.	Illusie, L.	Ulrich, B.
Traverso, C.	Józefiak, T.	Vázquez-Gallo, M. J.
Ulrich, B.	Johnsen, T.	van Gastel, L. J.
Welters, G. E.	Kempf, G.	van Straten, D.
Xambó, S.	Kim, B.	Vogel, W.
	Kleiman, S.	Welters, G. E.
	Krasiński, T.	Weyman, J.
	Laksov, D.	Winiarski, T.
	Lascoux, A.	Xambó, S.
	Lazarsfeld, R.	Zelevinsky, A.

Table 6. Authors in the core components.

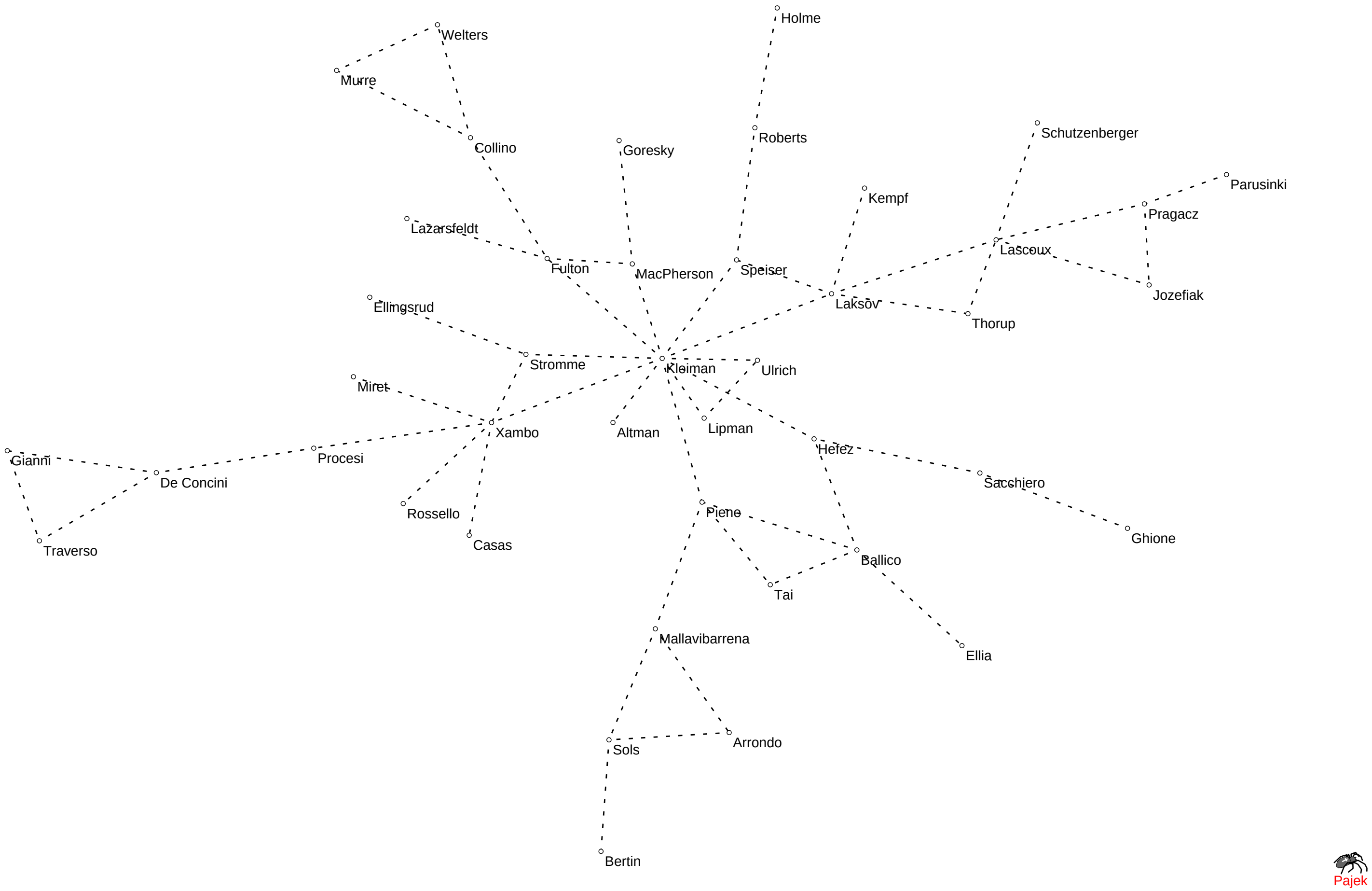

Fig. 4. The core component of the enumerative geometry world in 1992.

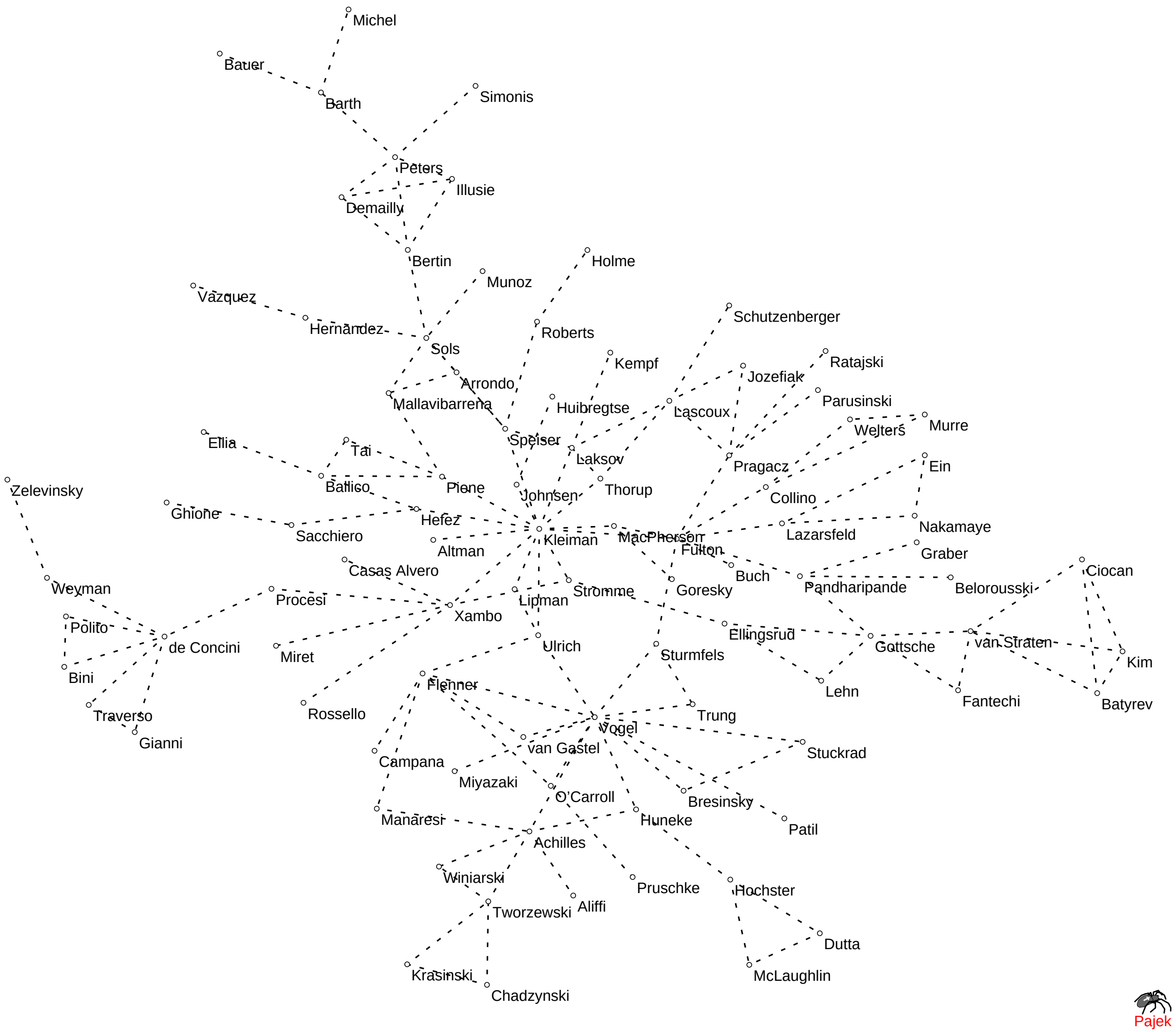

Fig. 5. The core component of the enumerative geometry world in 2002.