The cognitive structure of antibiotic resistance research, 2007-2016

2017-03-24

Bo Jarneving

Digital Services

at Gothenburg University Library

Table of contents

Abstract

On basis of 46,220 source papers from Web of Science, the cognitive structure of antibiotic resistance research during the period of 2007-2016 was assessed by bibliometric methods. Co-citation relations between the most cited references were applied as input to a cluster analysis which resulted in 66 clusters. The features of these clusters were analyzed with regard to size, internal cohesion, temporal features and impact. It was concluded that the analysis provided with a comprehensible depiction of the cognitive structure of the field of antibiotic resistance. It was suggested that the study may lay a foundation for further, more detailed explorations and an example of application was presented.

1. Introduction

This study is part of an ongoing project with the aim of providing bibliometric information to researchers associated with *Centre for Antibiotic Resistance Research* (CARe) at the University of Gothenburg. Various methods and data are tested with regard to applicability and utility. In this study, traditional and well known bibliometric methods are applied on a large set of source data collected from the Web of Science's citation indexes. Citation analysis provides with the most applied and developed set of methods in the field of bibliometrics, and co-citation analysis has been applied for several decades for science mapping purposes. In bibliometric jargon the objective is commonly to map the cognitive or intellectual structure of a field of knowledge. A field is mostly defined by a select set of highly used and cited journals and the cognitive structure corresponds to the sub-division of the field into research themes or specialties. In this study the point of departure in a pre-defined set of journals was not feasible on grounds of the multidisciplinary character of the field. Hence, queries directed to the topic-search field in the Web of Science interface had to substitute for the more common approach. The model applied in this study is coined the 'co-citation cluster model' and it can be concluded that besides the method of co-citation analysis, there is also a method of clustering. Cluster methods have been applied in both the sciences and social sciences for a long time and has been frequently applied for classification purposes. Classification in the current context should be understood as the identification of coherent research themes or specialties.

1.1 Statement of purpose

The purpose of this study was to identify and map specialties from the field of antibiotic resistance research published during the period 2007-2016 by means of citation analysis. A specialty is defined as a two-component construct consisting of a cluster of papers knit together by shared citing papers, the so called intellectual base, and a set of papers giving reference to papers in the clusters, the specialties current citing literature. The following research questions were stated:

1. How can the specialty structure be described in terms of research themes?
2. How can temporal features of the field and its research themes be described?
3. How can the impact on the research community by these research themes be described?

Also, it has been the intention to generate a report where results and supplementary data could be applied for further inquiry and elaborations. For this reason, Section 5.1 is dedicated to the illustration of how to use this information.

2. Method

2.1 Data

The queries 'antimicrobial resistance' and 'antibiotic resistance' were applied in order to retrieve papers on antibiotic resistance research published during the period 2007-2016. Taking the union of the retrieved sets of papers, a total of 46,220 research articles in English were downloaded. In total 700,589 distinct publications were cited by these source papers. The distribution of citations over cited

references was relatively even with a Gini-index of 54 %[1]. As we wish to map the more frequently cited references, we need to apply some threshold of selection. Such choices are made by rule of thumb and in this case it was decided that the minimum number of citations should be 100. This corresponded to 545 distinct cited references. This mode of selection implies that both early publications with a continuous, moderate citation rate as well as more recent and immediately recognized publications would be included.

Supplementary data are available and referred to at appropriate locations in the sequent sections.

2.2 Co-citation analysis

The co-citation analysis method is based on the assumption that there is an intellectual linkage between a citing and a cited paper as well as an intellectual linkage between two papers cited by the same third papers. When such associations are cumulated as the number of shared citing papers for a pair of cited papers, we arrive at the co-citation coupling strength. However, the co-citation coupling strength is not an optimal measure of association between two papers as the probability of being co-cited increases by the number of citations. For this reason, a common measure of association in this context is Salton's cosine formula[1] here applied as:

$$NCS_{ij} = \frac{C_{ij}}{\sqrt{C_i \cdot C_j}}$$

where

$NCS_{ij} =$ the normalized coupling strength between paper *i* and paper *j*

$C_{ij} =$ number of co-citations of paper *i* and paper *j*

$C_i =$ number of received citations for paper *i*

$C_j =$ number of recieved citations for paper *j*

In co-citation analysis, citation and co-citation (or normalized co-citation) thresholds are applied in order to separate signal from noise, hence only the more significant cited and co-cited papers are selected for the analysis.[3] the minimum number of citations was set to 100 whereas the normalized coupling strength was set to 0.1

2.3 The co-citation cluster model

The co-citation cluster model involves a method for the partitioning of a set of papers into disjoint groups, i.e. a cluster analytical method. Traditionally the single link method (nearest neighbor method) has been applied on grounds of being practical when handling large sets of data. In this study a variation of the single-link method implemented in the software Bibexcel has been applied for that same reason[2].

The concept of co-citation cluster involves two components: a set of highly cited and co-cited papers, the co-citation cluster or the cluster-core, and the much larger set of citing papers (the current citing literature) giving rise to the co-citations (Figure 1). It is assumed that the cluster-core represents the core of theories and methods, and that the current citing literature share research focus, theoretical

[1] The gini-index is a well known measure of eveness (equality-inequality) of a distribution, typically the eveness of the distribution of income for a population. In bibliometric resarch it is commonly used for the assessment of the concentration (inequality) of papers to some analyzed unit, for instance regions, countries or institutions. See alos Lorenz curves.

approach or method.[4] Hence, in the analysis, both a *citing pack* of papers as well as a *cited pack* are elaborated on.

Once the clusters have been generated, the identification of their subject contents is usually done by assessing the titles of the co-citing papers (the cited pack). The reason being that most of the bibliographic information of the co-cited papers is missing as cited references mostly are available only in an abbreviated form.[2] In this study, however, co-cited references were identified either in the *Web of Science* database or elsewhere on the Internet, and bibliographic data were studied in order to label each cluster. It should be emphasized that this labeling is accomplished on a layman level and that other interpretations of clusters' subject contents may be preferred by a field expert. In the supplementary data, Web of Science id:s and titles are available for the possible rectification of cluster-labels (topics).

Other useful information is based on statistics. These statistics will provide with insight into clusters' quality in a statistical sense and helps us decide whether a cluster should be considered 'real' or is only an artefact of the method. This is assessed in terms of *internal cohesion* and *external isolation*. For this assessment two corresponding measures (1) *the Average Coupling Strength for a Cluster C* ($ACS(C)$) and (2) *the Average Cluster strength between two Clusters C and C'* $\left(ACS(C, C')\right)$ are applied. They are defined as:

(1)

$$ACS(C) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} CS(d_i d_j)}{\binom{n}{2}}$$

where

n = the number of papers in a cluster C

CS = the number of co-citations for two papers d_i, d_j

and

$d_i d_j \ (\in C)$

(2)

$$ACS(C, C') = \frac{\sum_{i=1}^{k-1} \sum_{j=1}^{m} CS(d_i d_j)}{k \cdot m}$$

where

CS = the number of co-citations for two papers d_i, d_j

and

k and m are the sizes for cluster C and C'

and

$d_i \in C, d_j \in C'$.

[2] Cited references belonging to source papers are identified with a string of bibliographic data, for example *Emsley P, 2004, V60, P2126, ACTA CRYSTALLOGR D*, a cited reference belonging to Cluster 1. The inherent information is still sufficient for retrieving records applying the cited reference search function in *Web of Science*. As a matter of fact, nearly all cited references could be retrieved this way.

Conclusively, a cluster should have an internal coherence exceeding the strength of relation with any other cluster, that is, $ACS(C) > ACS(C, C')$ for a cluster *C*.

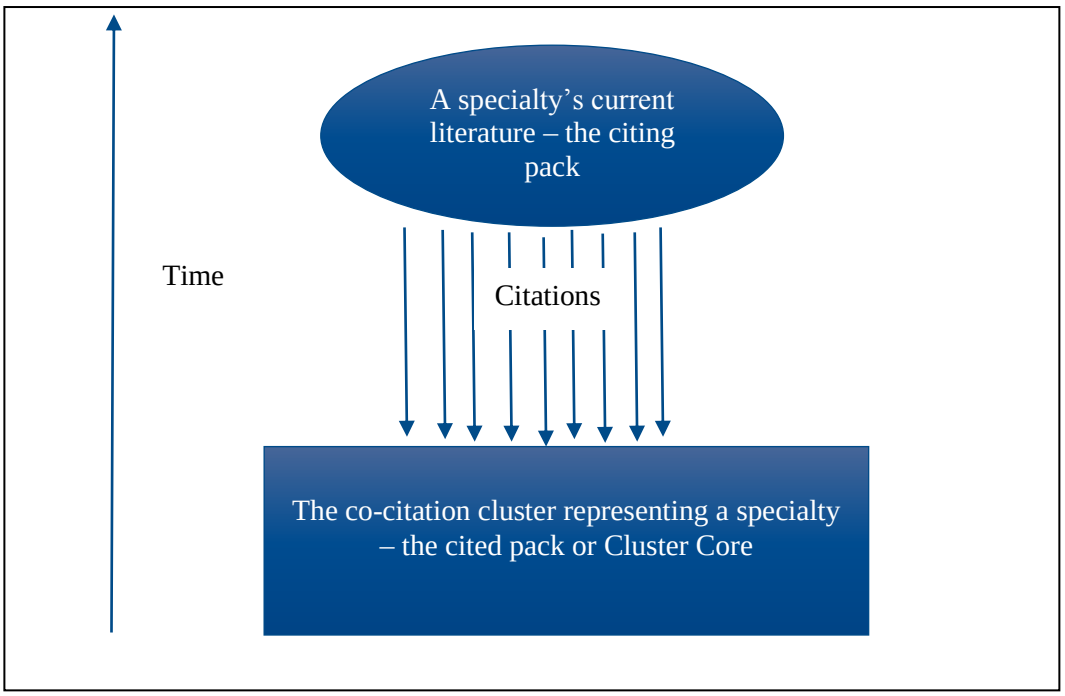

Figure 1. The Co-citation cluster model.

3. Findings

In this section, the research questions are elaborated on. First, let us reiterate these:

1. How can the specialty structure be described in terms of research themes?
2. How can temporal features of the field and its research themes be described?
3. How can the impact on the research community by these research themes be described?

The elaboration on research question 1 covers results from the clustering of papers in terms of size, internal coherence, labeling and the cognitive structure. The latter elaborates on the distribution of Journal Subject Categories, the relations between clusters as well as the relation between clusters and corresponding sets of citing papers. The following sections are dedicated to this research question:

- Section 3.1 *The cluster analysis*
- Section 3.2 *The cognitive structure*
- Section 3.3 *The current citing literature*

Research question 2, is elaborated by describing the growth of the field of antibiotic resistance research during the period of observation and by measuring the distance in time between the citing and the cited pack of a research theme, assessing the extent of continuous citation (viability) respectively more immediate recognition through citation. The Section 3.4 *Temporal aspects* deals with this research question.

With regard to research question 3, impact in terms of three different citation based indicators is analyzed in Section 3.5 *Impact*.

In order to facilitate an overview of information pertaining to the 66 identified clusters, the following variables are compiled in Appendix A:

- Label (Research theme)
- Number of papers
- Average coupling strength
- Average publishing year for the citing pack
- Average publishing year for the cited pack
- Distance in years between the citing and cited pack, i.e., the citing-cited distance
- Number of papers giving rise to at least one co-citation
- Number of citing papers
- Number of received citations
- Citations per paper
- Citations per paper divided by the distance in years

The same data are available in another format as supplementary data S6.

3.1 The cluster analysis

A total of 51,011 co-cited reference pairs with a co-citation strength between 214 and 1 were computed. After normalizing the raw co-citation frequency as described in the method section, a threshold of *NCS* = 0.1 was applied. This resulted in a reduction of the number of co-cited reference pairs to 1,446. Next, these data were applied to a clustering routine basically working as a ‘single-link’ algorithm, resulting in the generation of 66 co-citation clusters of varying sizes. The distribution of papers over clusters is presented in Table 1. The modal size was 3 (5 – a ‘tie’) and the average cluster size 6.6 (Md =6). Cf. supplementary data S1.

Table 1. The distribution of papers over number of clusters.

# Papers	Frequency
3	11
5	11
7	10
4	9
6	9
8	4
10	4
9	2
20	2
11	1
13	1
15	1
21	1
Sum	**66**

The cluster quality, i.e., the extent to which a cluster is coherent was measured by the *Average Coupling Strength*, as described in the method section. The arithmetic mean for the *ACS* was 27, the median 23 and the distribution positively skewed (Figure 2). Cf. supplementary data S2.

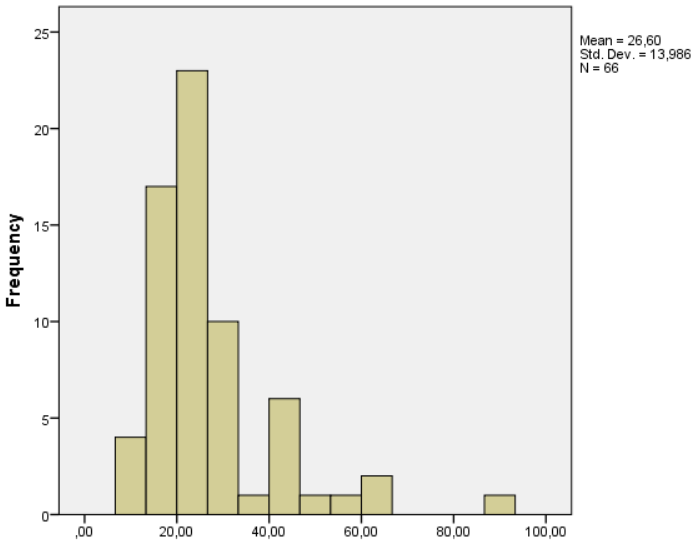

Figure 2. The distribution of ACS(C) over 66 clusters.

The next issue was to decide on the cognitive content of clusters. Here titles, abstract texts and key words were studied. As these data were not immediately available, each cited reference had to be looked up in *Web of Science*, which makes this approach impractical for larger scale studies. In Table 2, all 66 cluster labels are given. As can be noted, the labeling resulted frequently in label-names based on species or family. This reflects the semantic content of the titles, but other facets such as diagnosis, therapy or epidemiology may not be reflected. An approach combining MESH-terms with Web of Science data is currently being developed and would complement with other aspects.

Table 2. Labeling of 66 clusters on basis of titles from the cited packs.

Cluster	Label	Cluster	Label
1	Molecular Graphics I	34	Software for describing microbial communities
2	Molecular Graphics II	35	Biological cost of antibiotic resistance
3	Management of Helicobacter pylori I	36	Multiple sequence alignment analysis tools
4	Management of Helicobacter pylori II	37	Antibiotics and cell death
5	Resistant Neisseria gonorrhoeae	38	Carbapenemases
6	Antimicrobial peptides	39	Integrons
7	Enterococcus virulence determinants	40	Antimicrobial resistance genes of Escherichia coli
8	Persister cells and tolerance	41	Antimicrobial consumption and resistance
9	Pathogenic Escherichia coli	42	Antibiotics and environment II
10	Silver nanoparticles as antimicrobial agent	43	Aminoglycosides
11	Ciprofloxacin and Ceftazidime resistance	44	Outer membrane permeability
12	Methicillin-resistant Staphylococcus aureus	45	Escherichia coli K-12 genes
13	Mechanisms of resistance to quinolones	46	Development of a bacterial biofilm
14	Genome sequence of resistant Staphylococcus aureus	47	Detection of Beta-Lactamase genes
15	Carbapenem resistance in Acinetobacter baumannii	48	(MLS) antibiotics and resistance
16	Lipid A modification	49	(MLS) antibiotics and resistance II
17	(Waste) water and resistance genes	50	Salmonella
18	Streptococcus pneumonia	51	Pseudomonas aeruginosa and cystic fibrosis
19	Acinetobacter baumannii: Emergence and epidemiology	52	Read alignments
20	Escherichia coli resistance strains	53	Staphylococcus aureus
21	Antibiotic resistance genes	54	Beta–Lactamases structure and classification
22	Antibiotic resistance in lactic acid bacteria	55	Efflux–mediated drug resistance
23	New Metallo-beta-Lactamase Gene	56	Extended-spectrum beta lactamases II
24	Tetracycline resistance	57	Nosocomial infections
25	Integrons & gene cassettes	58	Gene studies and gene replacement (Pseudomonas aeruginosa)
26	Resistance in Acinetobacter baumannii strains	59	Gene transfer between bacteria
27	Bacterial biofilms	60	Pseudomonas aeruginosa and resistance
28	Methicillin-resistant staphylococcus aureus and communities	61	Genome annotation and sequencing
29	Extended-spectrum beta-lactamases	62	Campylobacter infections and food producing animals
30	Antimicrobial treatment of critically ill patients	63	Multidrug-resistance gram negative bacteria
31	Antibiotics and environment I	64	Antibiotic-resistant infections and community I
32	Urinary tract infections	65	Antibiotic-resistant infections and community II
33	Identification of plasmids	66	Antimicrobial susceptibility test

3.2 The cognitive structure

In the bibliometric literature the term ‘cognitive structure’ denotes the structure arrived at when partitioning a research field (mostly defined by its core-journals) into specialties or research-themes by means of quantitative methods. Some statistics is commonly used in order to map clusters’ quality and other features. The most immediate way of getting an over-view of the subject structure is to study a classification codes commonly provided by professional indexers. Our data, however, is collected from the multidisciplinary citation database *Web of Science* and there is no available typology on the paper-level. However, on the journal level, Web of Science provides with the so called *Journal*

Subject Categories. Each journal indexed in the Web of Science is assigned to one or several Journal Subject Categories. In praxis, all papers of a journal is assigned to the corresponding classification(s). The subject-classification is generally recognized as a neuralgic point and shortcomings of this particular classification scheme have been pointed out, though no real substitute has been presented as yet. Nevertheless, this classification scheme has been in use for a long period of time and is well known. We start by presenting the most frequent journal subject categories derived from the original set of 46,220 downloaded bibliographic records (Table 3). In total 128 distinct Journal Subject Categories were identified where frequencies ranged from 1 to 15603. As can be appreciated, *Microbiology* is ranked number one with a considerable gap to the next rank – *Infectious diseases*. The next notable gap is between *Pharmacology & Pharmacy* and *Biotechnology & Applied Microbiology*. Most notably, the three first categories retrieve 49 % of all distinct papers. The next notable gap between frequencies is seen at rank 13 where 84 % of all papers are retrieved. Conclusively there is a strong concentration of papers to a few categories. Cf. supplementary data S3.

Table 3. The distribution of Journal Subject Categories from the original set of source papers where N = 46,220. The 15 most frequent categories are shown. The total number of categories was 128.

Rank	Frequency	Journal Subject Category
1	15603	Microbiology
2	8357	Infectious Diseases
3	8251	Pharmacology & Pharmacy
4	4065	Biotechnology & Applied Microbiology
5	3815	Biochemistry & Molecular Biology
6	3092	Science & Technology - Other Topics
7	2623	Immunology
8	2515	Chemistry
9	2297	Veterinary Sciences
10	2266	Food Science & Technology
11	1837	General & Internal Medicine
12	1828	Environmental Sciences & Ecology
13	1587	Public, Environmental & Occupational Health
14	989	Agriculture
14	979	Research & Experimental Medicine

As mentioned, a cluster should optimally be internally coherent as well as clearly demarcated with regard to other clusters. This means that the *ACS(C)* within a cluster should be notably higher than the $ACS(C, C')$ to any other cluster. A relatively strong coupling to another cluster may reflect a similar research theme and that both clusters may be joined to one. Accordingly we applied a method for the visualization of the relations between clusters, applying the $ACS(C, C')$ as a measure of assocation. But first we need to fix the relation between internal coherence and external isolation. That is, at what strength should we regard a strong association between two clusters an indication of the split up of a research theme? As a point of reference it seems reasonable to use the arithmetic mean of the *ACS(C),* which was 27. Consulting the distribution of $ACS(C, C')$, only one cluster pair (Cluster 3 – Cluster 4) exceeds this threshold. Hence, the cluster solution seems generally quite valid. The arithmetic mean of the $ACS(C, C')$ was 1.1 but a range of 36 indicates relatively strong relations between some clusters (Figure 3). Cf. supplementary data S4.

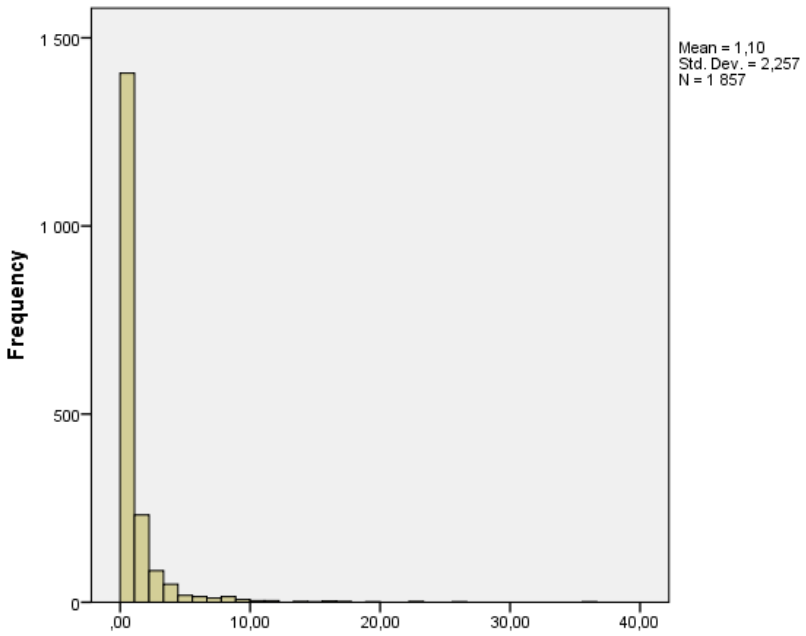

Figure 3. The distribution of the $ACS(C, C')$ over 1,857 cluster-pairs.

3.2.1 Bibliometric maps

Applying to the 95th percentile in the distribution of the $ACS(C, C')$ as the next threshold, clusters with a coupling strength > 9.8 to at least one other cluster were selected for mapping. Using Pajek and the Kamada- Kawai algorithm we arrive at the graph in Figure 4. Interpreting the configuration of the map, the underlying assumption is that the distance between clusters based on the $ACS(C, C')$ mirrors subject relatedness between clusters. A note of caution should be given: though bibliometric maps of this type have the capability of compressing data and visualizing patterns otherwise not accessible, the goodness of fit is almost never 100 %, hence, though the general pattern usually is comprehensible, on a detail level and for single cases, the distance may not perfectly correspond to the value of measured association. Conclusively, the configuration arrived at is a best compromise possible given the data and applied functions. This makes it a good idea to consult underlying data on different levels if one has a special focus of interest.

Relying on the assigned labels, it was tried to make sense of the configuration by discussing the rationale for the vicinity between clusters. Starting at the upper right quadrant, we notice the cluster-pair Cluster 3 and Cluster 4 which reflects research on *the management of Helicobacter Pylori*. This cluster-group is constituted by current papers on a short distance from the cluster core (cited pack). Both clusters have a strong internal cohesion.

Moving to the left of the map, the next cluster group is made up by three clusters: Cluster 46, Cluster 27 and Cluster 8 and gathers research themes involved with the *Development of a bacterial biofilm, Bacterial biofilms* and *Persister cells and tolerance*. The subject relatedness between these clusters is clear though the internal coherence is rather low. The distance between citing and cited packs is around a decade.

Next, Cluster 12 and Cluster 28 form a pair focusing on *Methicillin-resistant staphylococcus aureus.* Notably, the citing-cited distance is considerably smaller for Cluster 28, indicating an interest in a somewhat earlier literature. The internal coherence for cluster 12 is below the average.

Moving to the right and downwards, three clusters (19, 26 and 15) involved with research on *Acinetobacter Baumannii* are closely grouped together. Cluster 15 has a coherence well above the average but Cluster 26 and Cluster 19 are less coherent. All three clusters are based on a short citing-citation relationship and current research is cited.

A bit more to the right side of the map Cluster 44 and Cluster 16 form a pair connecting research on *Outer membrane permeability* with research on *Lipid A modification*. The subject relatedness rests on the relation between lipid A and 'outer membrane'. Cluster 44 is internally less coherent while Cluster 16 is above the average: Both clusters have a citing-cited distance of 10 years.

Immediately below this group Cluster 1 and Cluster 2 are located near one another. Both clusters deal with *Molecular Graphics* and are strongly internally coherent. The difference between these clusters is the distance to the cluster core. For Cluster 1 the citing-cited distance is 14 years while for Cluster 2 it is only five years.

Mowing upwards and to the left, another pair is seen. Cluster 64 and Cluster 65 connect on grounds of a common focus on *antibiotic-resistant infections and community*. Their internal coherence is somewhat below the average, but both cluster have a small citing-cited distance.

Now we move to the left from the previously discussed pair of clusters to a group of five clusters: Cluster 47, Cluster 29, Cluster 20, Cluster 9 and Cluster 33. In order to get a clearer picture of the cluster we zoom in on the relations between these particular clusters (Figure 5). We note that this sub-graph is in fact a complete graph where every node is connected with every *n*-1 other node, hence a complete mutual relationship is at hand. Considering titles, there seems to be a clear connection between *Detection of Beta-Lactamase genes* (Cluster 47) and *Extended-spectrum beta-lactamases* (Cluster 29) and between *Escherichia coli resistance strains* (Cluster 20) and *Pathogenic Escherichia coli* (Cluster 9). Cluster 33 *Identification of plasmids* has its strongest connection with cluster 20. Cluster 20 is the most novel cluster with a short citing-cited distance and two clusters, Cluster 47 and Cluster 33, have an internal cohesion lower than average.

Mowing downwards to the cluster-pair Cluster 11 *(Ciprofloxacin and Ceftazidime resistance)* and Cluster 13 (*Mechanisms of resistance to quinolones*), research on quinolones and resistance is reflected. Both clusters have citing-cited distance below the average and Cluster 13 is just below the average *ACS(C).*

To the left from this cluster-pair we see Cluster 25 (*Integrons & gene cassettes*) and Cluster 39 (*Integrons*). The internal cohesion is somewhat above the average for Cluster 25 and somewhat below the average for Cluster 39. The citing-cited distance is 12 respectively 10 years.

A cluster-group of four members can be seen at the left-lower quadrant of the cluster map. This group is focused on the influence of environmental factors on antibiotic resistance: *Antibiotics and environment I* (Cluster 31), *(Waste) water and resistance genes* (Cluster 17), *Antibiotics and environment II* (Cluster 42) and *Tetracycline resistance* (Cluster 24). The citing-cited distance varies between 5 and 11 years and all but Cluster 24 are below the mean *ACS.*

Cluster 38 (*Carbapenemases*) and Cluster 23 (*New Metallo-beta Lactamase-gene*) are situated in the center of the graph, representing research on beta lactamase from various angles. The larger Cluster 38 mirrors epidemiology, outcomes and detection of beta-lactamase infections while the smaller Cluster 23 focus on NDM-1 resistance. Both clusters are generated by short-distance citations. Cluster 38, shows some topic drift and has a relatively low internal coherence, while the small Cluster 23 is very coherent.

Finally, the cluster-pair Cluster 52 (*Read Alignments*) and Cluster 61 (*Genome annotation and sequencing*), situated to the left of the previous cluster pair, mirror research on bio-informatics. Both clusters are generated on a small citing-cited distance. Cluster 52 is the more coherent cluster.

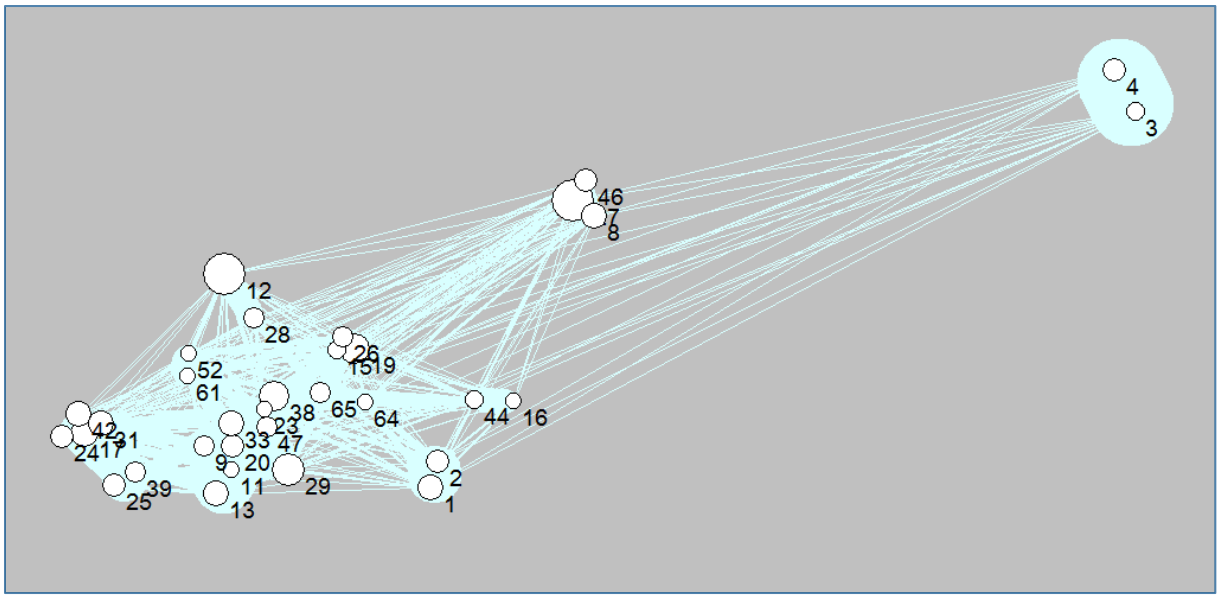

Figure 4. Co-citation cluster map based on 33 clusters with at least one link to another cluster above the threshold of coupling strength. The network analytical program Pajek was applied using the Kamada-Kawai algorithm. Circle sizes are proportional to number of cluster members while distances and width of lines correspond to the strength of association.

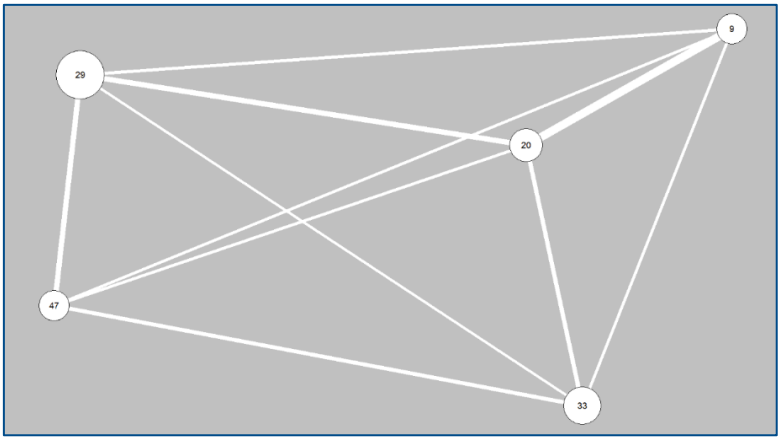

Figure 5. Zooming in on the configuration of five clusters from the graph in Figure 4. Circle sizes are proportional to number of cluster members while distances and width of lines correspond to the strength of association. The network analytical program Pajek was applied using the Kamada-Kawai algorithm.

Now we have a visualization of those clusters that may be considered closely connected and labels seem to be in line with statistical data. One should be generous with the use of bibliometric maps as a single map seldom covers all aspects. Hence, another map is presented in Figure 6 where the remaining 33 clusters are depicted. This map informs us about the cluster structure when only the more isolated clusters are mapped. We can appreciate a notable difference between this map and the map in Figure 4. This network is much sparser with longer distances between clusters reflecting weaker relations between clusters. Those cluster pairs that are connected near the threshold of coupling strength appears as pairs or small groups, for example *Cluster 53-Staphylococcus aureus* and *Cluster 14-Genome sequence of resistant staphylococcus aureus* connected at $ACS(C, C')$= 8.3. Most central of the map is *Cluster 21 Antibiotic resistance genes*. It is the largest cluster and its central position indicates connections with a large part of the other clusters. As the distance should mirror or in some way approximate the strength of cognitive relationship between clusters, we would expect that topics represented by clusters on a long distance from each other are quite different. For example, the label of Cluster 5 at the far end to the left in the map is *Resistant Neisseria gonorrhea* while the label

of Cluster 10 at the right far end of the map is *Silver nanoparticles as antimicrobial agent*. Clearly, one would not expect the simultaneous citation of papers from these two clusters and in fact no such exists in this data.

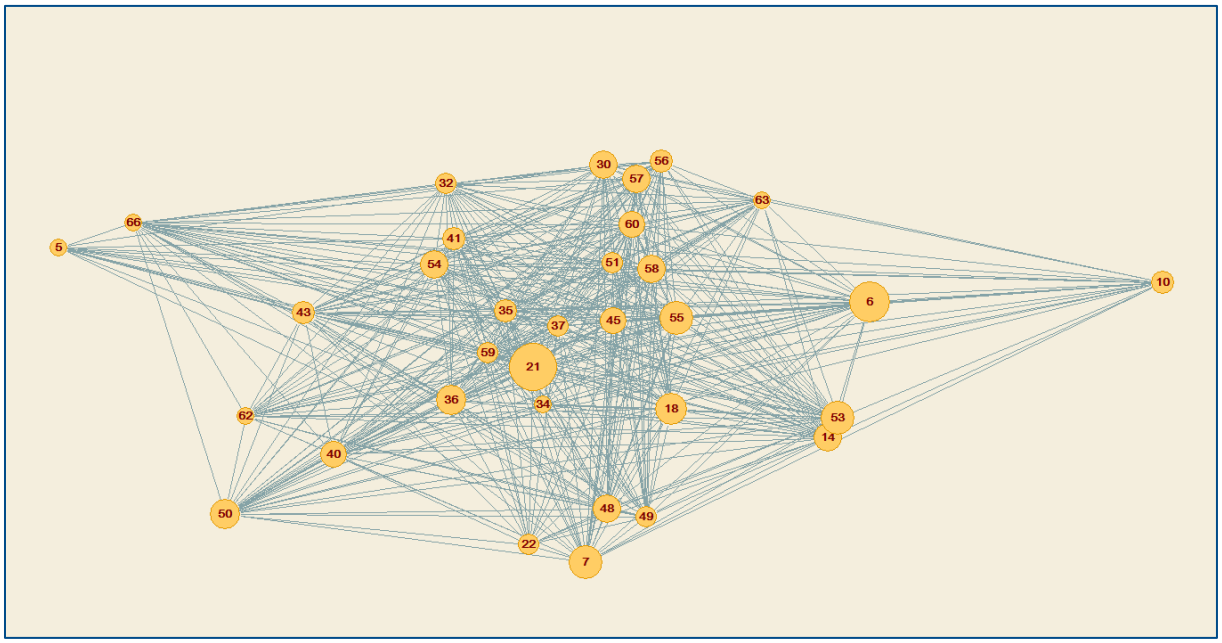

Figure 6. Co-citation cluster map based on the relations between 33 clusters connected below the threshold. Pajek was applied using the Kamada-Kawai algorithm. Circle sizes are proportional to number of cluster members while distances and width of lines correspond to the strength of association.

Cluster labels are listed below in order to facilitate the interpretation of the map:

Cluster	Label	Cluster	Label
5	Resistant Neisseria gonorrhoeae	45	Escherichia coli K-12 genes
6	Antimicrobial peptides	48	(MLS) antibiotics and resistance
7	Enterococcus virulence determinants	49	(MLS) antibiotics and resistance II
10	Silver nanoparticles as antimicrobial agent	50	Salmonella
14	Genome sequence of resistant Staphylococcus aureus	51	Pseudomonas aeruginosa and cystic fibrosis
18	Streptococcus pneumonia	53	Staphylococcus aureus
21	Antibiotic resistance genes	54	Beta–Lactamases structure and classification
22	Antibiotic resistance in lactic acid bacteria	55	Efflux–mediated drug resistance
30	Antimicrobial treatment of critically ill patients	56	Extended-spectrum beta lactamases II
32	Urinary tract infections	57	Nosocomial infections
34	Software for describing microbial communities	58	Gene studies and gene replacement (Pseudomonas aeruginosa)
35	Biological cost of antibiotic resistance	59	Gene transfer between bacteria
36	Multiple sequence alignment analysis tools	60	Pseudomonas aeruginosa and resistance
37	Antibiotics and cell death	62	Campylobacter infections and food producing animals
40	Antimicrobial resistance genes of Escherichia coli	63	Multidrug-resistance gram negative bacteria
41	Antimicrobial consumption and resistance	66	Antimicrobial susceptibility test
43	Aminoglycosides		

As aforementioned, the optimal cluster structure presents us with homogenous clusters with strong internal cohesions and considerably less strong external connections. We can appreciate that this holds on a general level as depicted by the two histograms (Figure 2 and Figure 3). One should bear in mind

that the best suited application of bibliometric mapping concerns the partitioning of a known subject-field into clearly demarcated specialties or sub-fields. In our case, the level of classification is not that clear as we deal with a multidisciplinary context.

Finally, we may apply all coupling links between all 66 clusters – regardless of strength –and zoom out in order to grasp the overall structure of the total graph (Figure 7). As can be appreciated, the approximate positions of clusters presented in Figure 4 are the same. The difference is that more clusters adhere to the graph and there is less of a clustering tendency in the map. In a sense, we could justify the delimitation of number of clusters to 46 by merging clusters in accordance with the information in the graph in Figure 4 (33 original clusters merged to 13 larger clusters), however, at the same time some information would be lost. Summing up, the cluster-cocitation map does not present us with clearly demarcated groups and clusters are rather evenly distributed. There is more of a center-periphery pattern with dissimlar groups separated by distance.

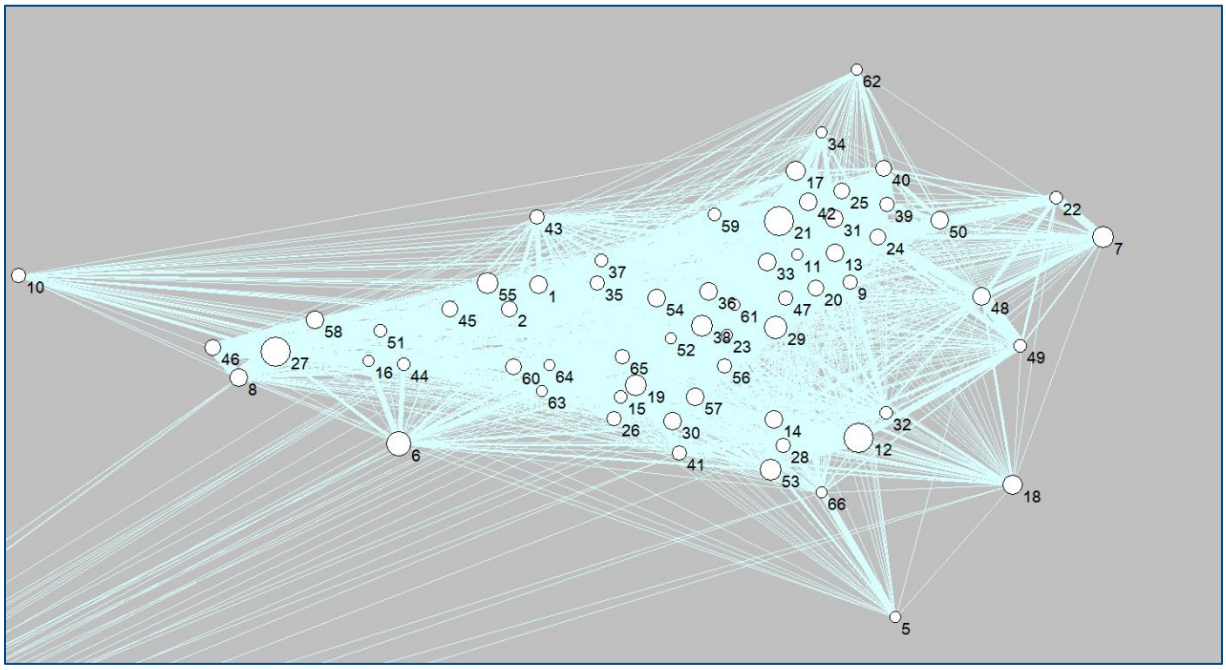

Figure 7. Map of all relations between 66 clusters. Cluster 3 and Cluster 4 are outside of the frame (left lower corner) in order to enhance the readability of the graph. Circle sizes are proportional to number of cluster members while distances and width of lines correspond to the strength of association. Cluster labels are listed below in order to facilitate the interpretation of the map:

Cluster	Label	Cluster	Label
1	Molecular Graphics I	34	Software for describing microbial communities
2	Molecular Graphics II	35	Biological cost of antibiotic resistance
3	Management of Helicobacter pylori I	36	Multiple sequence alignment analysis tools
4	Management of Helicobacter pylori II	37	Antibiotics and cell death
5	Resistant Neisseria gonorrhoeae	38	Carbapenemases
6	Antimicrobial peptides	39	Integrons
7	Enterococcus virulence determinants	40	Antimicrobial resistance genes of Escherichia coli
8	Persister cells and tolerance	41	Antimicrobial consumption and resistance
9	Pathogenic Escherichia coli	42	Antibiotics and environment II
10	Silver nanoparticles as antimicrobial agent	43	Aminoglycosides
11	Ciprofloxacin and Ceftazidime resistance	44	Outer membrane permeability
12	Methicillin-resistant Staphylococcus aureus	45	Escherichia coli K-12 genes

13	Mechanisms of resistance to quinolones	46	Development of a bacterial biofilm
14	Genome sequence of resistant Staphylococcus aureus	47	Detection of Beta-Lactamase genes
15	Carbapenem resistance in Acinetobacter baumannii	48	(MLS) antibiotics and resistance I
16	Lipid A modification	49	(MLS) antibiotics and resistance II
17	(Waste) water and resistance genes	50	Salmonella
18	Streptococcus pneumonia	51	Pseudomonas aeruginosa and cystic fibrosis
19	Acinetobacter baumannii: Emergence and epidemiology	52	Read alignments
20	Escherichia coli resistance strains	53	Staphylococcus aureus
21	Antibiotic resistance genes	54	Beta–Lactamases structure and classification
22	Antibiotic resistance in lactic acid bacteria	55	Efflux–mediated drug resistance
23	New Metallo-beta-Lactamase Gene	56	Extended-spectrum beta lactamases II
24	Tetracycline resistance	57	Nosocomial infections
25	Integrons & gene cassettes	58	Gene studies and gene replacement (Pseudomonas aeruginosa)
26	Resistance in Acinetobacter baumannii strains	59	Gene transfer between bacteria
27	Bacterial biofilms	60	Pseudomonas aeruginosa and resistance
28	Methicillin-resistant staphylococcus aureus and communities	61	Genome annotation and sequencing
29	Extended-spectrum beta-lactamases	62	Campylobacter infections and food producing animals
30	Antimicrobial treatment of critically ill patients	63	Multidrug-resistance gram negative bacteria
31	Antibiotics and environment I	64	Antibiotic-resistant infections and community I
32	Urinary tract infections	65	Antibiotic-resistant infections and community II
33	Identification of plasmids	66	Antimicrobial susceptibility test

3.3 The current citing literature

Focusing on the other component of the co-citation cluster construct, a specialty's current citing literature, a much larger and more diverse collection of papers with more topic spread is at hand. This larger set of current papers is a source for further elaboration and analysis. In particular, the topic drift in the set of citing and co-citing papers should be of particular interest, mirroring different uses of the earlier literature. However, comprehensible information is not readily at hand (should we not settle with serendipity and browsing) and we would need complementary analyses when zooming in on co-citation clusters' citing packs. Hence, the detailed analysis of each of the 66 clusters' citing packs cannot be covered in this study, though a mode of investigation will be suggested in Section 5.1.

Still, we would like to have some idea about the subject content of the more diverse citing side. Identifying those papers that exclusively co-cite one particular cluster – so called central papers – we assume that we find the more appropriate representatives of a cluster's subject content.[3] On the average, 26 percent of a cluster's citing papers are central papers, hence this set of papers is a select and smaller part of all papers citing a cluster. Furthermore, considering the length of the reference lists, the *relative citation frequency* (number of citations/number of references) to a cluster would better reflect the strength of relation between the citing paper and the cited cluster. By selecting papers with the highest relative citation frequency from the set of central papers, we form a sub-set representative of the subject content of the sited side. In Appendix B the best representatives (as defined) of clusters' citing packs are listed along with their corresponding cluster labels. In most cases there is an obvious subject relatedness between cluster label and selected exemplar-paper. In some cases, however, the trace of association is not clear. For instance the exemplar paper corresponding to Cluster 34 *Software for describing microbial communities* has the title *Evaluation of the nasal microbiota in slaughter-age pigs and the impact on nasal methicillin-resistant Staphylococcus aureus (MRSA) carriage* and the relationship is not obvious in a semantic sense. However, studying the abstract and reference list of the exemplar paper we find a clear relation between biometrics and

research on MRSA and food-animals. This is an example of the common case where a highly cited method paper connect a variety of empirical papers. Cf. supplementary data S5.

3.4 Temporal aspects

The distribution of papers over publication years is presented in Figure 8 and as can be concluded, there is a considerable increase in number of papers over time. Knowing the total number of papers on the topic (*N*=72399) and the number of papers indexed before 2007 (*n* = 26179) we can compute the annual growth rate for the period as: $(72399/26179)^{(1/10)} - 1 = 0.107$. Hence, this literature grows by 11 percent annually, which is clearly above that for science in general.[6] One has to keep in mind though that there are large variations between fields.

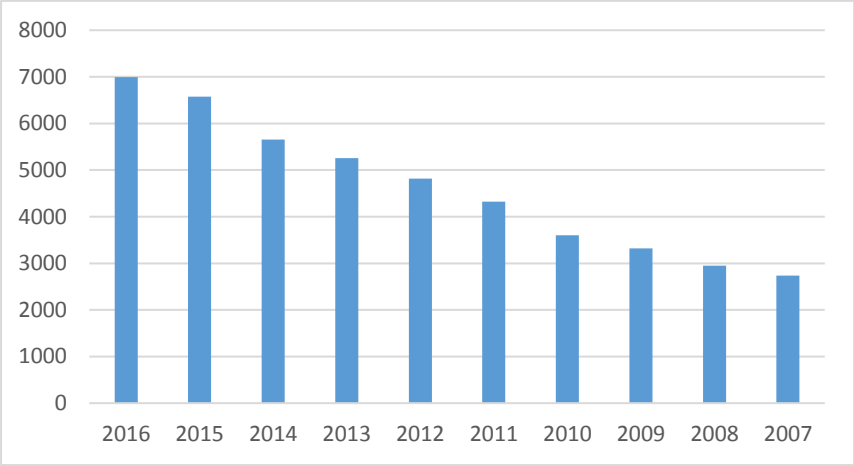

Figure 8. The distribution of research articles in English indexed in Web of Science and published during the period 2007-2016.

Considering research themes, on average, the distance between the citing and the cited pack of a cluster (citing-cited distance) was 8.4 years. The minimum was 2.1 years (Cluster 66 *Antimicrobial Susceptibility Test*) and the maximum 16.6 years. (Cluster 58 *Gene studies and gene replacement*). In the first case (Cluster 66), the small distance reflects updates of standards from the Clinical and Laboratory Standards Institute (CLSI). In the second case (Cluster 58), we can appreciate that the literature on 'Gene studies and gene replacement' is indeed viable and is still cited. Are there any notable trends? About half of all clusters have distances larger than the average and 18 clusters have a distance exceeding 10 years. The 10 most viable clusters are shown in Table 4.

Table 4. Ten clusters with the largest distance between the citing and the cited pack**.**

Cluster	Label	Distance
58	Gene studies and gene replacement (Pseudomonas aeruginosa)	17
1	Molecular Graphics I	14
54	Beta–Lactamases structure and classification	13
45	Escherichia coli K-12 genes	13
46	Development of a bacterial biofilm	12
9	Pathogenic Escherichia coli	12
57	Nosocomial infections	12
53	Staphylococcus aureus	12
48	(MLS) antibiotics and resistance	12
25	Integrons- gene cassettes	12

Considering clusters with relatively small distances between the citing and the cited packs, research themes characterized by more rapid integration of previous research are identified (Table 5).

Table 5. Ten clusters with the smallest distance between the citing and the cited pack.

Cluster	Label	Distance
6	Antimicrobial susceptibility test	2
3	Management of Helicobacter pylori I	3
5	Resistant Neisseria gonorrhoeae	3
23	New Metallo-beta-Lactamase Gene	4
37	Antibiotics and cell death	4
20	Escherichia coli resistance strains	4
21	Antibiotic resistance genes	5
64	Antibiotic-resistant infections and community I	5
61	Genome annotation and sequencing	5
65	Antibiotic-resistant infections and community II	5
2	Molecular Graphics II	5
17	(Waste) water and resistance genes	5

Interestingly, clusters with similar subject content may be separated by the time factor. For instance, papers in the clusters *Molecular Graphics I* and *Molecular Graphics II* share research focus but belong to different clusters and the most likely explanation for this separation is the time factor. These two clusters are, as one would expect, strongly connected above the mean $ACS(C)$. A similar example is the separation of papers from clusters Management of *Helicobacter pylori I* and Management of *Helicobacter pylori II*. In this case the distance between clusters considerably smaller (7 years), but one may suggest the same underlying cause for the separation of papers in different clusters. Consult Appendix A or supplementary data S6.

3.5 Impact

The relation between the citation frequency of a paper and its quality has been debated over time. In the present context it will, however, suffice with the presumption that a high citation frequency of a research paper mirrors the use of it in succeeding published research. Another term for this is impact. As previously elaborated on, there is a relation between citedness and publication date or age of a paper. Hence, though "raw" citation frequencies inform us about the citation volume as such, we need to relate citation impact with the age, should we want to compare papers. On the cluster level this was accomplished by dividing citations per paper with the distance between the citing and the cited pack (CPP/D). Computing the correlation coefficient (r) between CPP/D and CPP we arrive at $r = +0.43$. This result underlines the rationale of normalizing citation frequencies as the coefficient r should be considered low in this context. In Table 6, the 10 clusters with the highest score on CPP/D are presented. Consult Appendix A or supplementary data S6.

Table 6. 10 clusters with the highest score on CPP/D.

Cluster	Label	*D*	*CPP*	*CPP/D*
23	New Metallo-beta-Lactamase Gene	3.6	290.3	79.6
65	Antibiotic-resistant infections and community II	5.0	291.6	58.1
41	Antimicrobial consumption and resistance	9.5	226.0	23.9
6	Antimicrobial peptides	9.5	220.5	23.2
24	Tetracycline resistance	10.9	252.7	23.2
29	Extended-spectrum beta-lactamases	10.6	246.0	23.1
36	Multiple sequence alignment analysis tools	11.5	255.4	22.2

27	Bacterial biofilms	11.3	219.6	19.5
9	Pathogenic Escherichia coli	12.5	235.0	18.8
57	Nosocomial infections	12.4	212.4	17.1

Though the need for relative indicators have been (rightfully) emphasized in the bibliometric literature[6], raw frequencies or absolute values also contain useful information. The mere volume of citation mirrors the concentration of effort even if the impact is diluted by numer of papers or years. In table 7 the most cited clusters – regardless of publication date or number of cluster members – are listed, reflecting volumes of research communication for the period of observation. Consult Appendix A for more data or supplementary data S6.

Table 7. The ten most cited clusters.

Cluster	Label	Size	Average publishing year	No citing papers	No Citations
27	Bacterial biofilms	21	2001	2233	4611
12	Methicillin-resistant Staphylococcus aureus	20	2001	2426	4095
21	Antibiotic resistance genes	20	2009	2065	3483
6	Antimicrobial peptides	15	2003	1613	3308
29	Extended-spectrum beta-lactamases	13	2001	1889	3198
19	Acinetobacter baumannii: Emergence and epidemiology	10	2006	1256	2074
36	Multiple sequence alignment analysis tools	8	2001	1562	2043
38	Carbapenemases	11	2007	1183	1744
55	Efflux–mediated drug resistance	10	2003	1082	1535
33	Identification of plasmids	8	2001	1157	1522

4. Conclusions

The research on antibiotic resistance draws on knowledge from many different fields. In total 128 fields were involved as reflected by counting assigned journal subject categories. It was further shown that *Microbiology*, *Infectious Diseases and Pharmacology & Pharmacy* together cover for nearly half of all published papers. The cluster analysis resulted in the sub-division of the field in 66 clusters. Clusters were generally coherent and demarcated and the final map with all 66 clusters showed a somewhat even distribution of clusters and more of a center-periphery pattern than that of a grouping. The labeling of clusters revealed that the classification accomplished by the cluster analysis to a large extent gathered papers on basis of similar families or species.

The field showed a rapid growth during the period of observation and an annual increase by 11 percent. On the average, a research theme had a distance (as defined) of 8.4 years between its citing pack and its cited pack. The smallest distance of 2.1 years was assigned the research theme *Antimicrobial susceptibility test* and the longest of 16.6 years *Gene studies and gene replacement.* About half of all clusters had distances larger than the average.

The impact was measured by two citation based indicators which identified the most cited clusters both in terms of citation averages normalized by time and absolute frequencies. With regard to normalized citation averages, the research theme *New Metallo-beta-Lactamase Gene* was ranked first and in the case of absolute frequencies *Bacterial biofilms*.

5. Discussion

This study connects to early bibliometric traditions where the focus was more on scientific information provision than on research evaluation.[6, 7] The results arrived at give an intelligible over-view of the intellectual structure of the field of antibiotic resistance research. With a point of departure in this information, more specific and possibly useful analyses may be accomplished. Considering the validity of the results it should be emphasized that bibliometrics is no hard science and there is mostly an interval of variation in which different method applications may be adequate. For instance, selected thresholds may be varied as well as indicators and measures, and last but not the least, the algorithm of clustering. This implies that given a fixed objective of analysis, varying results may be accomplished. Thus, bibliometric mapping exercises should be appreciated as heuristic and explorative tools. With these delimitations before the mind, a multitude of explorative approaches suitable for analyzing the formal part of research communication is at hand. In the following, a mode of application will be presented that suggests and illustrates a more detailed analysis of one of the research themes identified in this study.

5.1 Applications

The point of departure is a presumed information need concerning *Antimicrobial peptides*, corresponding to Cluster 6. Considering the configurations in the co-citation maps we can appreciate that this cluster is isolated in a statistical sense with no strong connection to any other cluster. Hence, it would probably suffice to focus on this cluster only. We can conclude that the internal coherence is above the mean and that the citing-cited distance is ten years (cf. Appendix A). Thus we conclude that the cited literature, the cluster-core, is quite viable.

Considering both co-citation links and publication years, an interesting graph may be accomplished (Figure 8). According the width and number of connecting lines, the center of gravity in this graph is located at the paper by Zasloff published in Nature 2002 (*Antimicrobial peptides of multicellular organisms*). This paper has its two strongest relations to the paper by Brogden KA published in Nature Reviews Microbiology in 2005 (*Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria?)* and the paper by Hancock REW published in Nature Biotechnology in 2006 (*Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies*). Next, let us assume there is an implicit y-axis representing publication years in the graph and that all papers are ordered accordingly[5]. Reading the graph from top to bottom we can chronologically follow the development of research of this cluster, starting at 1997 and ending at 2012 and at the same time arrive at an understanding of the co-citation relations. In Table 8 we list all titles and publication years in accordance with the graph in Figure 8.

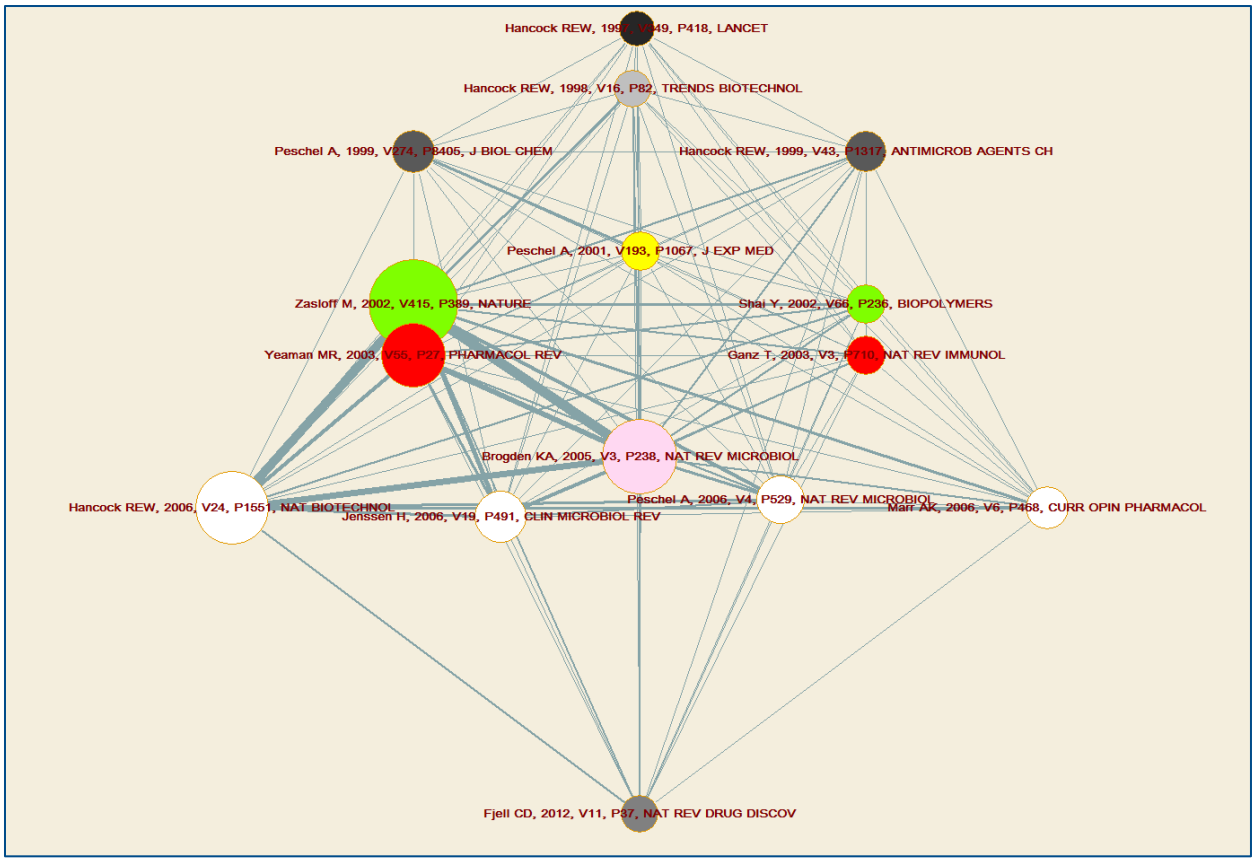

Figure 9. Graph of Cluster 6: the configuration of nodes representing clusters in the graph is arranged according to publication year (a presumed y-axis). Circle sizes are proportional to number of cluster members while length and width of lines correspond to the strength of association. Colors represent different publication years.

Table 8. Papers from Cluster 6 ordered ascending by publication year.

Year	First author	Title
1997	Hancock REW	Peptide antibiotics
1998	Hancock REW	Cationic peptides: a new source of antibiotics
1999	Peschel A	Inactivation of the dlt operon in Staphylococcus aureus confers sensitivity to defensins, protegrins, and other antimicrobial peptides
1999	Hancock REW	Peptide antibiotics
2001	Peschel A	Staphylococcus aureus resistance to human defensins and evasion of neutrophil killing via the novel virulence factor MprF is based on modification of membrane lipids with L-lysine
2002	Zasloff M	Antimicrobial peptides of multicellular organisms
2002	Shai Y	Mode of action of membrane active antimicrobial peptides
2003	Yeaman MR	Mechanisms of antimicrobial peptide action and resistance
2003	Ganz T	Defensins: Antimicrobial peptides of innate immunity
2005	Brogden KA	Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria?
2006	Hancock REW	Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies
2006	Peschel A	The co-evolution of host cationic antimicrobial peptides and microbial resistance
2006	Jenssen H	Peptide antimicrobial agents
2006	Marr AK	Antibacterial peptides for therapeutic use: obstacles and realistic outlook
2012	Fjell CD	Designing antimicrobial peptides: form follows function

Consulting Appendix A we conclude that 1613 papers have cited at least one of the papers in this cluster. Basically, we acknowledge that there is some cognitive linkage between each of these citing

papers and the cluster and that all citing papers are potentially useful. However, one may assume more topic drift in the citing pack and that not all papers are of equal interest to us. Though browsing and serendipity may open doors for new discoveries, we would anyway try to delimit the amount of information we need to assess. One way is to select citing papers with the highest relative citation frequency (Table 9). We can appreciate that the highest relative citation frequency is 27 %. Conclusively, we would concentrate on the higher ranks when expanding/exploring this research theme as originally represented by the cluster core. Besides applying the principles of analysis described above, it should be suggested that simplistic approaches like varying the sort order of Appendix A facilitates the identificstion of top-clusters for several variables like impact (CPP/D), size, number of citing papers etc.

Table 9. Papers citing Cluster 6 ordered by ascending publication year (PY) and then descending by relative citation frequency. The top 32 papers according to relative citation frequency.

PY	Title	Citations	Relative citation fr.
2007	Improved antimicrobial peptides based on acyl-lysine oligomers	4	0,27
2007	Design of antimicrobial compounds based on peptide structures	2	0,20
2007	Application of the Suzuki-Miyaura cross-coupling to increase antimicrobial potency generates promising novel antibacterials	5	0,19
2007	Peptide fraction inhibiting plant pathogen growth predominated in cell wall extracts from young plants or in soluble cell fraction from expanded leaves from eggplants	2	0,18
2007	Antimicrobial activity of rationally designed amino terminal modified peptides	4	0,17
2007	Impairment of innate immune killing mechanisms by bacteriostatic antibiotics	4	0,17
2008	The GraRS regulatory system controls Staphylococcus aureus susceptibility to antimicrobial host defenses	5	0,19
2008	Pleurocidin-derived antifungal peptides with selective membrane-disruption effect	5	0,19
2008	Design and synthesis of cationic antimicrobial peptides with improved activity and selectivity against Vibrio spp	3	0,17
2008	Structure-activity relations of parasin I, a histone H2A-derived antimicrobial peptide	5	0,17
2010	Interaction of cationic antimicrobial peptides with phospholipid vesicles and their antibacterial activity	5	0,18
2010	Synergy with Rifampin and Kanamycin Enhances Potency, Kill Kinetics, and Selectivity of De Novo-Designed Antimicrobial Peptides	5	0,16
2011	Structure-activity relationship of buffalo antibacterial hepcidin analogs	4	0,17
2012	Antitumor effects and cell selectivity of temporin-1CEa, an antimicrobial peptide from the skin secretions of the Chinese brown frog (Rana chensinensis)	5	0,22
2012	Therapeutic antimicrobial peptides may compromise natural immunity	2	0,17
2012	Subacute toxicity of antimicrobial peptide S-thanatin in ICR mice	4	0,15
2013	Non hemolytic short peptidomimetics as a new class of potent and broad-spectrum antimicrobial agents	2	0,25
2013	Antimicrobial peptide from spider venom inhibits Chlamydia trachomatis infection at an early stage	5	0,20
2013	pH Dependence of Microbe Sterilization by Cationic Antimicrobial Peptides	8	0,15
2014	Antimicrobial activity of de novo designed cationic peptides against multi-resistant clinical isolates	3	0,23
2014	De Novo Design and Their Antimicrobial Activity of Stapled Amphipathic Helices of Heptapeptides	4	0,18
2014	Expression of Recombinant Carcinoembryonic-Antigen-Human-beta-Defensin-2 Gene in Colon Cancer HCT116 Cells	2	0,18
2014	In vitro determination of the short-chain synthetic peptide RP 13 antimicrobial activity	6	0,16
2014	Antimicrobial Effects of a Hexapetide KCM21 against Pseudomonas syringae pv. tomato DC3000 and Clavibacter michiganensis subsp michiganensis	5	0,16
2014	Balteatide: A Novel Antimicrobial Decapeptide from the Skin Secretion of the Purple-Sided Leaf Frog, Phyllomedusa baltea	4	0,15

2014	High Cell Selectivity and Low-Level Antibacterial Resistance of Designed Amphiphilic Peptide G(IIKK)(3)I-NH2	6	0,15
2015	Antimicrobial peptides: has their time arrived?	4	0,22
2015	N-Capping Effects of Stapled Heptapeptides on Antimicrobial and Hemolytic Activities	4	0,18
2015	Membrane Interactions of Synthetic Peptides with Antimicrobial Potential: Effect of Electrostatic Interactions and Amphiphilicity	6	0,16
2016	Antimicrobial and Hemolytic Activity of Stapled Heptapeptide Dimers	5	0,26
2016	Short cationic lipopeptides as effective antibacterial agents: Design, physicochemical properties and biological evaluation	5	0,18
2016	A novel antimicrobial peptide derived from membrane-proximal external region of human immunodeficiency virus type 1	5	0,16

References

1. Salton, G., McGill, M.J. Introduction to modern information retrieval. Auckland: McGraw-Hill; 1983.
2. Åström, F. et al. Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at his 60th Birthday. In Special volume of the e-zine of the ISSI 05-S June 2009.
3. Brahm R.R., Moed H.F, and van Raan F.J. Mapping Science by Combined Co-citation and Word Analysis. I. Structural Aspects. Scientometrics 42(4):223-251, 1991.
4. Franklin J.J. and Johnston R. Co-citation Bibliometric Modelling and R & D Management. In Handbook of Quantitative Studies of Science and Technology. Editor van Raan A.F.J. 1988.
5. Power-point presentation by Olle Persson, European Summer School in Scientometrics (ESSS), 2012.
6. Lutz Bornmann, L., Rüdiger, M. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the American Society for Information Science and Technology, 66(11):2215–2222, 2015.
7. Schubert, A., Glänzel, W., and Braun, T. Against absolute methods: Relative scientometric indicators and relational charts as evaluation tools. In Handbook of Quantitative Studies of Science and Technology. Editor van Raan A.F.J. 1988.
8. Kessler, M.M. An experimental study of bibliographic coupling between technical papers. Massachusetts Institute for Technology, Lincoln Laboratory, 1962.
9. Griffith, B et al. The structure of scientific literatures II: toward a macro- and microstructure for science, Science studies, 4:339-365, 1974.

Appendix A

Cluster variables for 66 co-citation clusters

A: Research theme

B: Number of papers

C: Average coupling strength

D: Average publishing year for the citing pack

E: Average publishing year for the cited pack

F: Distance in years between the citing and cited pack, i.e., the citing-cited distance

G: Number of papers giving rise to at least one co-citation

H: Number of citing papers

I: Number of received citations

J: Citations per paper

K: Citations per paper divided by the distance in years (J/F)

Cluster	A	B	C	D	E	F	G	H	I	J	J/F
1	Molecular Graphics I	7	55	2012	1998	14	378	671	1373	196	14
2	Molecular Graphics II	6	43	2014	2009	5	217	434	830	138	27
3	Management of Helicobacter pylori I	4	64	2014	2011	3	193	374	654	164	55
4	Management of Helicobacter pylori II	6	47	2012	2006	7	308	581	1054	176	26
5	Resistant Neisseria gonorrhoeae	3	64	2014	2011	3	116	217	371	124	41
6	Antimicrobial peptides	15	30	2013	2003	10	858	1613	3308	221	23
7	Enterococcus virulence determinants	10	19	2012	2001	12	316	854	1386	139	12
8	Persister cells and tolerance	7	30	2013	2004	9	215	642	1023	146	17
9	Pathogenic Escherichia coli	5	43	2013	2000	12	240	850	1175	235	19
10	Silver nanoparticles as antimicrobial agent	5	38	2014	2005	9	163	377	634	127	15
11	Ciprofloxacin and Ceftazidime resistance	3	42	2013	2006	6	103	331	446	149	23
12	Methicillin-resistant Staphylococcus aureus	20	19	2012	2001	11	747	2426	4095	205	18
13	Mechanisms of resistance to quinolones	8	26	2012	2005	7	267	807	1259	157	22
14	Genome sequence of resistant Staphylococcus aureus	7	23	2012	2004	8	150	765	1037	148	19
15	Carbapenem resistance in Acinetobacter baumannii	4	46	2013	2006	7	149	416	625	156	23
16	Lipid A modification	3	31	2012	2002	10	68	253	333	111	11
17	(Waste) water and resistance genes	9	25	2014	2008	5	294	712	1233	137	26
18	Streptococcus pneuminiae	9	18	2012	2003	8	285	807	1238	138	16

19	Acinetobacter baumannii: Emergence and epidemiology	10	29	2013	2006	7	484	1256	2074	207	30
20	Escherichia coli resistance strains	6	31	2013	2009	4	204	635	949	158	35
21	Antibiotic resistance genes	20	15	2014	2009	5	709	2065	3483	174	39
22	Antibiotic resistance in lactic acid bacteria	4	29	2012	2004	8	95	334	464	116	14
23	A new Metallo-beta-Lactamase Gene	3	87	2014	2010	4	171	655	871	290	80
24	Tetracycline resistance	6	43	2012	2001	11	292	1076	1516	253	23
25	Integrons & gene cassettes	6	30	2012	2001	12	191	553	856	143	12
26	Resistance in Acinetobacter baumannii strains	5	20	2013	2007	6	115	437	590	118	20
27	Bacterial biofilms	21	23	2013	2001	11	1147	2233	4611	220	19
28	Methicillin-resistant staphylococcus aureus and communities	5	27	2012	2006	6	178	761	982	196	32
29	Extended-spectrum beta-lactamases	13	29	2012	2001	11	752	1889	3198	246	23
30	Antimicrobial treatment of critically ill patients	7	21	2012	2002	10	249	996	1334	191	18
31	Antibiotics and environment I	7	22	2013	2005	8	245	904	1243	178	23
32	Urinary tract infections	4	23	2012	2004	9	105	456	576	144	17
33	Identification of plasmids	8	16	2013	2001	11	296	1157	1522	190	17
34	Software for describing microbial communities	3	23	2014	2009	6	62	254	319	106	19
35	Biological cost of antibiotic resistance	5	30	2013	2004	8	175	623	855	171	20
36	Multiple sequence alignment analysis tools	8	21	2013	2001	11	383	1562	2043	255	22
37	Antibiotics and cell death	4	23	2014	2010	4	109	626	749	187	46
38	Carbapenemases	11	15	2013	2007	6	387	1183	1744	159	28
39	Integrons	5	24	2011	2001	10	145	579	769	154	15
40	Antimicrobial resistance genes of Escherichia coli	6	17	2012	2003	8	133	631	817	136	16
41	Antimicrobial consumption and resistance	5	27	2012	2003	9	201	898	1130	226	24
42	Antibiotics and environment II	7	24	2013	2005	8	262	828	1199	171	23
43	Aminoglycosides	5	20	2012	2002	10	121	447	604	121	12
44	Outer membrane permeability	4	21	2013	2003	10	88	473	580	145	15
45	Escherichia coli K-12 genes	5	24	2012	2000	13	245	848	1149	192	15
46	Development of a bacterial biofilm	6	14	2012	2000	12	123	571	732	122	10
47	Detection of Beta-Lactamase genes	5	22	2013	2006	7	162	776	964	193	26
48	(MLS) antibiotics and resistance I	7	11	2012	2000	12	173	859	1056	151	13

49	(MLS) antibiotics and resistance II	4	16	2012	2002	10	66	416	497	124	12
50	Salmonella	8	15	2013	2005	8	236	997	1315	164	21
51	Pseudomonas aeruginosa and cystic fibrosis	4	22	2012	2001	11	73	410	511	128	12
52	Read alignments	3	40	2014	2009	6	95	369	477	159	27
53	Staphylococcus aureus	10	10	2012	2000	12	256	1147	1477	148	12
54	Beta–Lactamases structure and classification	7	21	2013	2000	13	252	931	1263	180	14
55	Efflux–mediated drug resistance	10	14	2012	2003	9	311	1082	1535	154	16
56	Extended-spectrum beta lactamases II	5	12	2012	2006	6	77	496	593	119	18
57	Nosocomial infections	7	16	2012	2000	12	238	1199	1487	212	17
58	Gene studies and gene replacement (Pseudomonas aeruginosa)	7	14	2012	1995	17	194	738	977	140	8
59	Gene transfer between bacteria	4	15	2013	2004	9	63	411	487	122	13
60	Pseudomonas aeruginosa and resistance	6	15	2013	2005	7	152	720	908	151	21
61	Genome annotation and sequencing	3	26	2015	2010	5	68	417	490	163	35
62	Campylobacter infections and food producing animals	3	20	2011	2002	9	55	307	365	122	13
63	Multidrug-resistance gram negative bacteria	3	18	2012	2006	6	38	400	446	149	24
64	Antibiotic-resistant infections and community I	3	23	2014	2009	5	62	446	511	170	37
65	Antibiotic-resistant infections and community II	5	25	2013	2008	5	180	1246	1458	292	58
66	Antimicrobial susceptibility test	3	11	2015	2013	2	32	329	361	120	58

Appendix B

Exemplar papers: central papers with the highest relative number of citations to a particular cluster.

Cluster	Label	Docnr.	Exemplar paper
1	Molecular Graphics I	33879	Crystallization and preliminary crystallographic studies of the butyrolactone autoregula Streptomyces virginiae
2	Molecular Graphics II	45796	Crystal Structure of Escherichia coli originated MCR-1, a phosphoethanolamine transfer
3	Management of Helicobacter pylori I	1116	Antimicrobial Susceptibility-Guided Therapy Versus Empirical Concomitant Therapy for in a Region with High Rate of Clarithromycin Resistance
4	Management of Helicobacter pylori II	33739	A new 24 h ELISA culture based method for Helicobacter pylori chemosusceptibility
5	Resistant Neisseria gonorrhoeae	25302	Enhanced gonococcal antimicrobial surveillance in the era of ceftriaxone resistance: a re detection of the Neisseria gonorrhoeae H041 strain
6	Antimicrobial peptides	43953	Improved antimicrobial peptides based on acyl-lysine oligomers
7	Enterococcus virulence determinants	19073	Severe sepsis caused by a linezolid-resistant Enterococcus faecium in a 10-year-old girl a
8	Persister cells and tolerance	45287	Antibiotic treatment in vitro of phenotypically tolerant bacterial populations
9	Pathogenic Escherichia coli	6945	Antimicrobial Resistance and Virulence Characterization among Escherichia coli Clinical Infections in Pregnant Women
10	Silver nanoparticles as antimicrobial agent	8596	Antimicrobial effect of silver zinc oxide (Ag-ZnO) nanocomposite particles
11	Ciprofloxacin and Ceftazidime resistance	30007	Resistance to ampicillin, third-generation cephalosporins, ciprofloxacin, co-trimoxazole isolates of Salmonella enterica subsp enterica from Germany: Real problem or sporadic
12	Methicillin-resistant Staphylococcus aureus	14202	Molecular characterization of methicillin-resistant Staphylococcus aureus resistant to ti in a hospital in Brazil
13	Mechanisms of resistance to quinolones	15434	Plasmid-mediated quinolone resistance in typhoidal Salmonellae: A preliminary report f
14	Genome sequence of resistant Staphylococcus aureus	35040	Low-shear modelled microgravity alters expression of virulence determinants of Staphy
15	Carbapenem resistance in Acinetobacter baumannii	44851	Differences in phenotypic and genotypic traits against antimicrobial agents between Ac Acinetobacter genomic species 13TU
16	Lipd A modification	35252	Identification of a Functionally Important Loop in Salmonella typhimurium ArnT
17	(Waste) water and resistance genes	12394	Removal performance and changes in the microbial communities of SBRs under aerobic tetracycline pressure
18	Streptococcus pneuminiae	37898	Impact of pneumococcal conjugate vaccine on infections caused by antibiotic-resistant
19	Acinetobacter baumannii: Emergence and epidemiology	35044	Antibiotic resistance determinants in Acinetobacter spp and clinical outcomes in patien facility
20	Escherichia coli resistance strains	21346	Complete Sequence of pJIE186-2, a Plasmid Carrying Multiple Virulence Factors from a S coli O25 Strain
21	Antibiotic resistance genes	13409	Safely Coupling Livestock and Crop Production Systems: How Rapidly Do Antibiotic Resis following a Commercial Application of Swine or Dairy Manure?
22	Antibiotic resistance in lactic acid bacteria	31427	Antibiotic Resistance of Some Lactobacilli Isolated from the Digestive Tract of Broiler in
23	New Metallo-beta-Lactamase Gene	32100	First Identification of a Patient Colonized With Klebsiella pneumoniae Carrying bla(NDM
24	Tetracycline resistance	30453	Impact of Minocycline Ointment for Periodontal Treatment of Oral Bacteria
25	Integrons & gene cassettes	29543	Transcription of integron-harboured gene cassette impacts integration efficiency in clas
26	Resistance in Acinetobacter baumannii strains	29940	Genome Sequence of a Dominant, Multidrug-Resistant Acinetobacter baumannii Strain,
27	Bacterial biofilms	2264	Equine endometritis and biofilm forming Escherichia coli
28	Methicillin-resistant staphylococcus aureus and communities	37206	Risk Factors for Methicillin-Resistant Staphylococcal aureus Skin and Soft Tissue Infectio South Texas Ambulatory Research Network (STARNet) Study
29	Extended-spectrum beta-lactamases	33722	Rapid detection of TEM, SHV and CTX-M extended-spectrum beta-lactamases in Enterob mediated amplification with microarray analysis
30	Antimicrobial treatment of	33938	Resistant Pathogens in Nonnosocomial Pneumonia and Respiratory Failure Is It Time To

	critically ill patients		care-Associated Pneumonia?
31	Antibiotics and environment I	12495	Bacteriophages as a reservoir of extended-spectrum beta-lactamase and fluoroquinolon environment
32	Urinary tract infections	5049	A study of urinary tract infections due to multidrug resistant bacteria in critical care unit
33	Identification of plasmids	28600	Multilocus sequence typing of IncN plasmids
34	Software for describing microbial communities	14052	Evaluation of the nasal microbiota in slaughter-age pigs and the impact on nasal methic aureus (MRSA) carriage
35	Biological cost of antibiotic resistance	41604	Spread of a low-fitness drug-resistant Mycobacterium tuberculosis strain in a setting of virus prevalence
36	Multiple sequence alignment analysis tools	11589	Dysgonomonas macrotermitis sp. nov., isolated from the hindgut of a fungus-growing te
37	Antibiotics and cell death	22530	Antibiotic plasma levels in dogs with Otitis externa treated routinely with various topica
38	Carbapenemases	13139	Carbapenem-resistant enterobacteriaceae: analyzing knowledge and practice in healthc
39	Integrons	39659	Novel cassette arrays of integrons in clinical strains of Enterobacteriaceae in China
40	Antimicrobial resistance genes of Escherichia coli	36554	Distribution of Sulfonamide Resistance Genes in Escherichia coli and Salmonella Isolates Abattoirs in Ontario and Quebec, Canada
41	Antimicrobial consumption and resistance	44579	Global development of resistance - secondary publication
42	Antibiotics and environment II	16319	Antimicrobial resistance to 14 antimicrobials in marine coastal waters around Northern Resistance Index as a marker of ecological status
43	Aminoglycosides	15871	Aminoglycoside Resistance in Clinical Klebsiella pneumoniae Isolates
44	Outer membrane permeability	22616	Thymus maroccanus essential oil, a membranotropic compound active on Gram-negativ
45	Escherichia coli K-12 genes	40217	Antibiotic marker modifications of lambda Red and FLP helper plasmids, pKD46 and pCP chromosomal genes using PCR products in multidrug-resistant strains
46	Development of a bacterial biofilm	35760	Virulence and antimicrobial resistance determinants of human pathogenic and commen aeruginosa
47	Detection of Beta-Lactamase genes	29528	CMY-42, a Novel Plasmid-Mediated CMY-2 Variant AmpC Beta-Lactamase
48	(MLS) antibiotics and resistance	26443	Detection of the Major Macrolide Resistance Genes in Streptococcus suis Serotype 2 Isc
49	(MLS) antibiotics and resistance II	42636	Agar dilution method for detection of inducible clindamycin resistance in Staphylococcu
50	Salmonella	9681	Similarities between Salmonella Enteritidis isolated from humans and captive wild anim
51	Pseudomonas aeruginosa and cystic fibrosis	38959	Antimicrobial resistance in cystic fibrosis isolates of Haemophilus influenzae
52	Read alignments	24071	Complete Genome Sequence of the Strong Mutator Salmonella enterica subsp enterica
53	Staphylococcus aureus	38381	Reduced Susceptibility to Vancomycin Influences Pathogenicity in Staphylococcus aureu
54	Beta–Lactamases structure and classification	15463	Identification of Family Specific Fingerprints in beta-Lactamase Families
55	Efflux–mediated drug resistance	2466	A liposomal method for evaluation of inhibitors of H+-coupled multidrug transporters
56	Extended-spectrum beta lactamases II	10837	Rapid Detection of Extended-Spectrum-beta-Lactamase-Producing Enterobacteriaceae ESBL NDP Test
57	Nosocomial infections	21698	Predictors and outcomes of linezolid-resistant vancomycin-resistant Enterococcus: A ca
58	Gene studies and gene replacement (Pseudomonas aeruginosa)	32415	Identification of Pseudomonas aeruginosa genes associated with antibiotic susceptibilit
59	Gene transfer between bacteria	34286	The Influence of Horizontal Gene Transfer on the Mean Fitness of Unicellular Populatio
60	Pseudomonas aeruginosa and resistance	30307	Imipenem and ciprofloxacin consumption as factors associated with high incidence rate aeruginosa in hospitals in northern France
61	Genome annotation and sequencing	33016	Genome sequencing and annotation of a Campylobacter coli strain isolated from milk w
62	Campylobacter infections and food producing animals	35662	Antimicrobial resistance of thermophilic Campylobacter spp. isolated from cattle in Pola
63	Multidrug-resistance gram negative bacteria	36147	Pharmacokinetics of novel antimicrobial cationic peptides NAB 7061 and NAB 739 in rat administration
64	Antibiotic-resistant infections	27105	Tackling antibiotic resistance

	and community I		
65	Antibiotic-resistant infections and community II	31766	Neonatologists' Perceptions of Antimicrobial Resistance and Stewardship in Neonatal In
66	Antimicrobial susceptibility test	19781	Spectrum and potency of ceftaroline tested against leading pathogens causing skin and (2010)