Visualizing patent statistics by means of social network analysis tools[1]

Christian Sternitzke[2,1,2], Adam Bartkowski[1], Reinhard Schramm[1]

[1]Technische Universität Ilmenau, PATON – Landespatentzentrum Thüringen, PF 100 565, 98684 Ilmenau, Germany

[2]Universität Bremen, Forschungsgruppe Innovation und Kompetenztransfer, Wilhelm-Herbst-Strasse 12, 28359 Bremen, Germany

Abstract:

The present paper reviews the literature on social network analysis with applications to bibliometric data, and in particular, patent information. Several approaches of network analysis are conducted in the field of optoelectronics to exemplify the power of network analysis tools. Cooperation networks between inventors and applicants are illustrated, emphasizing bibliometric measures such as activity, citation frequency, etc. as well as network theoretical measures, e.g. centrality, betweenness, etc. In this context it is found that inventors who serve as interfaces or links between different inventor groups apply for technologically broader patents, hence, benefiting from their access to different knowledge through their position. Furthermore, citation networks of patent documents as well as patent applicants were drawn. Here, patent thickets could be identified. The position of applicants within citation networks seems to be useful in explaining behaviour of the applicants in the marketplace, such as cooperation or patent infringement trials.

Key words: patent statistics, cooperation, social network analysis, citation analysis, citation networks, optoelectronics, competitor analysis

*This is the accepted version of the paper. The final version is available under* https://doi.org/10.1016/j.wpi.2007.08.003

---

[1] An earlier version of this paper was presented at the PATINFO 2007 conference, June 14-15 in Ilmenau/Germany.

[2] Corresponding author: christian.sternitzke@tu-ilmenau.de, fax: +49 3677 694538

## 1. Introduction

In recent years, developments in the field of social network analysis brought up several software tools that facilitate visualization, analysis and interpretation of cooperation and citation data, explaining the relationship between technology fields (IPC classes), patent applicants, inventors, patent documents, etc. The tools furthermore allow the combination of several other types of analysis presented in an earlier paper [1] and enhance their visualization. The present paper, in contrast, highlights some types of network analysis of patent data which extend methodologies currently deployed in practice. First, these types of network analysis allow the identification of important players in technology fields or corporations. Second, their connectedness can be used in competitor analysis or for identifying partners for joint development projects. Third, the methodology also allows the identification of key patents, and fourth, rivalry between players in the technology field. The paper is organized as follows: Section 2 provides an overview of the literature on social network analysis for bibliometric data. Section 3 explains the methodology and technology field. Results and discussions can be found in section 4, conclusions follow.

## 2. Social network analysis of patent and literature information

Social network analysis explores the relationship ("ties", "arcs" or "edges") between actors ("nodes" or "vertices"). Historically, the methodology was focused on the relationship between humans. However, since the underlying algorithms originate from the field of graph theory and are universally applicable, modelling of technical relationships such as traffic over the internet also became popular (see e.g. [2]). When taking patent or literature information as a basis, nodes can represent individuals such as inventors, patent applicants, or documents like patents or scientific papers. Ties can symbolize cooperation between the nodes or citation links.

In the literature to date, very few studies employed social network analysis to more thoroughly investigate and visualize information originating from patent and literature data. Some of these studies used co-citation data, bibliographic coupling or even composite indicators serving as similarity measures (see e.g. [3-5]). Such relationships are frequently visualized by means of multivariate statistical methodologies, e.g. multidimensional scaling (MDS), and hence, are not discussed in the present paper focusing solely on methodologies which are frequently applied in social network analysis. Other studies relied on cooperation and citation data from publications contained in the Science Citation Index (SCI), such as research collaborations between corporations [6] or countries [7]. Another survey investigated medical research trajectories based on important publications from the SCI as well as cooperation behaviour between countries and research institutions [8]. An interesting approach from the patent analysis perspective analysed to what extent the density of patent citation networks was able to identify patent thickets. The MPEG patent pool, comprising all patents relating to this audio/video standard, served as an example, and it could be shown that the network density within the pool/thicket was higher than in surrounding areas [9].

The latter four studies all have in common that they only investigated two-dimensional data. In this case, two-dimensional means that only relationships between nodes (dimension 1) and ties (dimension 2) are illustrated. In general, software tools for social network analysis allow analysing multi-dimensional data. An example of multi-dimensional analysis was given by [10]. The cooperation behaviour of patent applicants and inventors is investigated, and the visualization of ties and nodes is enhanced with additional data, such as the technology field under consideration, the frequency of citations made or received, etc.

The position of individuals within corporate inventor networks is the subject of analysis as well, even though the data was not visualized. It was found that inventors who serve as interfaces or links between different inventor groups or R&D departments show a higher patent output [11] and citation frequency [12], implying that individuals who are positioned as information brokers between groups with different information backgrounds benefit from information flows and that this has a positive influence on their quantitative and qualitative output. Centrality within a network is also associated with a higher citation frequency of these individuals [12].

So far, social network analysis has only begun to invade the field of patent analysis. There already exist a number of commercial tools for patent analysis that, in particular, allow graphing of cooperation between inventors or applicants. Matheo Analyzer, Vantage Point, or Thomson Data Analyzer are prominent examples. These ready-to-use tools help to gain valuable insights into relationships in fields of search. Users, however, have more flexibility with tools from social network analysis, but coming at the price of investing slightly more time in data preparation.

## 3. Methodology and area of research

Methodologies deployed in this paper are static in nature, i.e. they represent snapshots at certain points in time and are retrospective over a period of time for a certain technology field. Dynamic investigations would be snapshots at many points in time which can be used to track e.g. technological developments. This paper performs static investigations on two different levels. On the first level, cooperation between (i) inventors and (ii) patent applicants is investigated, where cooperation between inventors is measured from co-inventorship of patent families and between patent applicants from co-application. In both cases, nodes represent inventors or applicants. On the second level, citation networks are investigated to demonstrate the relationship between (iii) patent families, and (iv) patent applicants.

The investigations on the first level allow identification of important players with many patents or citations, occupying key positions within a technology field. More important here, network analysis directly allows identification of the connectedness of these individuals within their (technological) environment. This means that individuals can, for instance, be recognized as hubs in a cooperation or citation network, or rather as bridges between different subnets. Identifying cooperation strength between nodes is another issue here. Taking these domains of information together, a better picture of the competencies of an individual can be created and, hence, used to create e.g. more promising researcher teams.

On the second level, key patents, characterized by a high citation frequency within a network, and their relationship to other patents can be identified more easily. It was showed that such analyses can be deployed to identify patent thickets [9]. When inventors or patent applicants are considered as nodes within a network, it can be demonstrated to what extent they build upon each others knowledge. Closeness between two nodes in a network signifies that they are technologically related. If, for instance, two applicants are situated closely together, and they do not cooperate, then they should be engaged in a high level of technological competition. However, if they cooperate, then it seems rather that they jointly develop new technology, using complementary competencies. To calculate the citation ties between applicants to better assess technological competitiveness between these players, two approaches are chosen:

i) Simple counts, i.e. if applicant A cited six patent families from applicant B, then six citations are counted.

ii) Multiple citation counts, i.e. if applicant A cited six patent families from applicant B, but if each patent family was cited twice (e.g. from two different patent families of applicant A) then twelve citations are counted. This approach should deliver more exact results than the former.

The field of study of this paper is a technology closely related to that in an earlier paper [1]. It is the field of III-nitrides, i.e. semiconductor materials that are particularly useful in optoelectronics, but receive more and more attention in the fields of power electronics, sensors, etc. The field is narrowed to patent families relating to light emitting diodes (LEDs) and laser diodes (LDs). Therefore, the field is broader than the one used in the previous analysis [1], in particular with respect to the spectrum of light emission and designated countries of the applications. Furthermore, the analysis is carried out without time limit, using the family-oriented Derwent World Patents Index (WPINDEX) database via STN International. Hence, early work in this field is also considered, even though major activities did not start until the early 1990s. In total, 4,753 patent families could be identified.[3] In order to obtain data for creating a citation network, results were transferred to Derwent Patents Citation Index (DPCI) database, which is complementary to WPINDEX and includes information on citations to and from patent families. This led to a subset of 2,631 patent families, significantly reducing the size of the dataset.

During the course of analysis, results are processed with the tool PATONanalist [1], creating cooperation and citation matrices to be imported into a social network analysis tool. However, applicant and inventor data had to be cleaned first: the general problem with such bibliometric data is type-I and type-II errors. The former are synonyms, i.e. a person appears under two different names within the database, mainly due to typographical errors. The latter error occurs in the case of homonyms, i.e. one spelling of a name stands for several individuals. The probability that type-II errors exist increases with the popularity of the name and the size of the technological field. Databases such as WPINDEX or SCI enforce type-II errors since they only include initials of individuals' first names.[4]

The search in WPINDEX identified 4,726 different names. Since the visualization of such a large dataset has a negative impact on the readability of the graphs, the analysis is limited to the most active inventors (as well as applicants). Therefore, the names of the most active inventors are manually screened for type-I errors and eventually merged. Based on this dataset, the top five percent of the inventors regarding patenting activity are selected. This threshold level yields 240 inventors who hold at least twelve patents.

---

[3] The patent search was conducted in early March, 2007.

[4] A possibility to reduce type-II errors would be linking the results from WPINDEX/DPCI to the INPADOC from the European Patent Office since the latter database provides full names of inventors. However, in some cases first and last names of inventors are interchanged, making this work a tedious task that was omitted for this paper which goal it is to provide an overview on the methodologies rather than the results of social network analysis for patent data.

Nakamura as the most active inventor holds 143 patent families. The WPINDEX raw data yields 2,187 patent applicants. Applicants are limited to institutional ones, type-I errors are treated as for inventors, and the threshold level is set to five percent as well. In total, 107 institutional patent applicants can be identified holding at least six patents. Among them, Toyoda Gosei is the most active one with 325 patent families.

The transfer of the search results from WPINDEX to DPCI results in a substantial loss of information. In total, only 2,792 different inventors and 1,732 different applicants are identified in the raw data. Applying the same data treatment procedures as for the WPINDEX data, 171 inventors can be identified who hold at least seven patents. Here, the most active inventor, Shibata, holds 32 patent families. So the majority of Nakamura's patents – and certainly the majority of the patents of many other inventors - do not seem to be included in the DPCI database. Regarding institutional patent applicants, the five percent threshold level yields 77 different applicants who hold at least six patents. Toyoda Gosei as the most active applicant had registered 215 patent families. Mergers and acquisitions between different patent applicants are not taken into account either for the WPINDEX or the DPCI dataset.

The filtering function of DPCI is rooted in the philosophy of the database: only the most important countries worldwide are included and citation information originates from international search reports of EPO or PCT applications, or from examination reports drafted during the examination procedure. Therefore, the DPCI serves as a filter for more important inventions. When, however, a complete picture of activities within a technology field is the goal of an analysis, based on similarity measures between inventors, applicants, or patent documents, then a co-word analysis based on WPINDEX data should yield more exhaustive results than the filtered DPCI data.[5]

A further step while preparing the data for network analysis is shortening the names of applicants to assure readability of the node labels. PATONanalist is used to create matrices and attribute lists to be imported into the network analysis software. Attribute lists incorporate additional information on nodes such as age, citation frequency, etc. in order to visualize multiple dimensions of the network. One strength of the common social network analysis tools is their flexibility in creating such attribute lists.

5 Other advanced text mining techniques are also possible, such as n-grammes, natural language processing in its various forms, etc.

There are a variety of software tools for social network analysis (for a review see [13]). The studies cited above deployed two different network analysis tools. Some used Pajek [14], others UCINET [15]. The former is freeware for non-commercial use and allows a multitude of different analyses. The latter, according to [13] the most commonly employed tool, is not freely available and cannot perform as many analyses as Pajek, however, its capabilities are more than sufficient to perform the network analyses described in this paper. In addition, UCINET is more intuitive to handle. Therefore, the analyses in the present paper are conducted with UCINET and visualized with Netdraw, a complementary program for network drawing.

A prominent family of algorithms for network visualisation relates to so-called spring embedders. Their basic principle is to consider ties and their strengths as forces. The goal is to minimize forces within the network and reach equilibrium through repositioning the nodes [16]. Frequently, there is no single solution for the state of equilibrium, meaning that graphs drawn with spring embedders may look slightly different each time, even though the states of equilibrium are equal.

## 4. Results and discussion

First, results of the investigations on the first level are discussed, i.e. cooperation between inventors and applicants. Figure 1 demonstrates inventor cooperation networks based on WPINDEX data. The circle size represents the number of patent families (minimum twelve). Inventors who are not connected within the sample can be found in the upper left corner of the graph.

It can be seen that Akasaki is situated in a subnet that is positioned almost in the centre of the graph. He is not only strongly connected to a handful of colleagues, but also has many ties going to other inventor groups. Most of his work seems to originate from collaborations with other active inventors. Clearly fewer ties are associated with Nakamura, the most active inventor in the field. This becomes even more obvious when modelling the ties of these two top performing inventors as egonets. Egonets show ties between one central node and the nodes directly connected to this central node (the surrounding nodes) as well as the ties between these directly connected nodes. Figure 2 (a) presents the egonet of Nakamura, while Akasaki's egonet can be found in figure 2 (b).

*{insert figure 1 and 2 about here}*

These egonets reveal that in Akasaki's network the exchange of implicit knowledge may have played a major role, while this hardly seems to be the case for Nakamura. Knowledge required for his inventions seems to come from public sources or his own creativity. Looking back at Figure 1, more inventor groups can be identified. An eye catcher is the subnet of Wen, Yu, Tu et al who are strongly connected among each other. These inventors worked for Sanyuan Optoelectric, Canyuan Photoelectric, and Formosa Epitaxy, respectively, and have been active in the technology field since 2004.

M. Yamada seems to play another interesting role since he serves as an interface or link between the groups of Akasaki (working primarily for Toyoda Gosei) and Nakamura (Nichia). Inventors "bridging" different research groups are interesting because they have easier access to knowledge from both groups. For the 240 inventors in Figure 1 we find a statistically significant relationship between serving as an interface or link ("bridge") and possessing more patents in technologically distinct IPC classes, a measure frequently chosen to describe the technological breadth of a patent ([17-19]). Network theory suggests a measure called "centrality betweenness" [20] for measuring to what extent a node serves as a bridge. Hence, the correlation of the number of different IPC classes (full class, 7-digit, and 4-digit) per patent and centrality betweenness is calculated for this inventor network of highly active individuals. The results (see Table 1) suggest that there is only a moderate but significant correlation between serving as a bridge in a network and the number of different IPC classes obtained.

*{insert Table 1 about here}*

Kidoguchi, situated "southeast" of Akasaki et al., only worked for Matsushita. Since he is connected to "geographically distant" inventor groups, he seems to have been involved in a variety of technologically distant projects (when assuming that different inventor groups work on different technologies) and hence, possess comprehensive knowledge in the field. In fact, Kidoguchi is one of the top ten inventors regarding the number of different full IPC classes. The inventor leading this ranking is Motoki who can be found in a relatively central position "north" of Akasaki. He has various ties to distant inventors. Mori comes second, localized "west" of Akasaki, and possesses about the same characteristics.

Coming back to Yamada, we searched for full names via the Esp@cenet database and reveal an obvious type-II error here: there are two inventors with the same initial but different first names. So not all bridges identified in Figure 1 are in fact bridges, and the results from Table 1 are biased.

Nevertheless, a closer examination of the relationship between spanning bridges and applying for more technologically different patents may yield interesting results.

Figure 3 presents the cooperation network between the most active institutional patent applicants in WPINDEX. Applicants not connected within the technology field are situated in the upper left corner. The graph shows what could already be expected from studying the inventor network of Akasaki et al.: Toyoda Gosei, a company Akasaki once worked for, is strongly connected in the field. It is striking that Sumitomo possesses a central role regarding cooperation with other active players, particularly Sony and Sharp. Further detailed analysis of patent texts might reveal that new technological areas were jointly developed. Also worth mentioning is the subnet were Lumileds is located in. Lumileds was founded as a joint venture between Philips and Agilent, bundling the optoelectronics business of both firms. Agilent, as a spin-off from Hewlett-Packard, was formed for a similar purpose. Recently, Philips acquired Agilent’s share in the business [21], so the majority of the subnet is now Philips.

Further insights are provided by the DPCI data. The filtering function of the database narrows the analysis to more important patents. Integrating the citation frequency allows identification of the most prominent inventors.

*{insert figure 3 and 4 about here}*

Since inventors are only required to hold at least seven patent families in the DPCI data (in comparison to twelve in the WPINDEX dataset), Figure 4 reveals more corporate subnets. One example for such a subnet is the group of inventors Doradzinski, Kanbara et al. from Ammono who obviously did contract R&D for Nichia [1]. Strauss, Han, Haerle et al. worked for Siemens/Osram, Edmond et al. for Cree, etc.. Hiramatsu, once a professor at Nagoya University, now Mie University, was active in a number of research projects with industrial partners like Toyoda Gosei, Mitsubishi, Sumitomo, and others. Being engaged with such a broad number of partners put him into a bridge-position between two larger subnets.

In the following, the second level of analysis, namely citation networks, will be discussed, beginning with citation networks of patents. However, drawing a network with all DPCI data, i.e. more than 2000 patent families, does not yield a readable result. Therefore, for illustrative purposes, the timeframe of the analysis is limited to 1990-1995 in order to show a citation network including patent families from many applicants. In total, 298 patent families were registered during this period. The sample is further limited to patent families originating from the most active patentees as already described in section 4, eliminating nine patent families. Only 180 of the remaining 289 patent families are connected via citation ties. The resulting network can be found in Figure 5. Node size is inversely proportional to patent age, i.e. older patents appear larger. Figure 5 illustrates that particularly Toyoda Gosei and cooperation partners as well as Nichia possess the most highly cited patent families in the sample, while Toyoda Gosei's patents are older than those from Nichia. A striking patent belongs to Rohm (JP08097470). This patent family (highlighted as egonet in Figure 6) cites as many as 80 patent families within the network, many belonging to Nichia. In total, this patent cites 229 other patent families as well as a multitude of non-patent references. The reason for so many backward citations is that the claims of the patent are particularly broad. Not surprisingly, in December 2000, Nichia sued Rohm for infringing its patents according to the database Litalert.

*{insert Figures 5 and 6 about here}*

In a further analysis, the whole DPCI dataset is drawn as a network (not illustrated in this paper). Among all 2631 patent families, 1336 are connected via 2671 ties (forward citations). Surprisingly, there are 5850 patent forward citations for all patent families, meaning that about 80 percent of all citations to these patent families come from outside the technology field.[6] Among these 1336 patent families, 211 do not originate from the most active patent applicants as mentioned in section 4. Hence, these highly active patent applicants are responsible for over 80 percent of all patent families. As mentioned before, a picture of the complete network is not easily readable. Therefore the dataset is further limited to five companies that entered cross-license agreements regarding white LEDs with Nichia, i.e. Toyoda Gosei, Lumileds (including Philips, Agilent and Hewlett-Packard), Cree, and Osram/Siemens [22]. Together, these companies hold 218 patent families connected via citation ties.[7]

Figure 7 demonstrates the patent citation network of these five patent applicants. The circle size is proportional to the number of citations received. It can be seen that, in particular, patent families from Nichia are highly cited, both within the network as well as in total. In addition, Nichia's patent families hold a central position in the network. Both aspects should strengthen Nichia's negotiation power in cross-license agreements.

[6] An explanation for this finding might be the already described relevancy of III-nitrides for a number of different technology fields such as sensors or power electronics, whereas the latter fields cite patents from optoelectronics that describe growth and structures (e.g. quantum wells) made of III-nitrides for the first time.

[7] Not all patents shown in Figure 7 relate to white LEDS.

Based on Figure 7, the five companies are analyzed regarding the density of their self-citation networks. While Nichia's self citation network is not very dense, the contrary holds true for Toyoda Gosei and Cree in particular (see Figure 8). It therefore seems that the latter company aims to protect its intellectual property by means of a company-specific patent thicket. There are some central patents, some highly cited, and a number of patent families clustered around, citing one or even more central patent families.

*{insert Figure 7 and 8 about here}*

In a next step, citation ties between applicants are modelled to better assess technological competitiveness between these players. Figures 9 and 10 present the results, based on the 77 most active applicants found in the DPCI database. Circle size of the nodes is determined by the total patent activity. At first glance, there seems to be hardly any difference between the two networks. A more detailed look reveals, for instance, that Sony is situated quite close to Nichia in Figure 9, but moves further away when taking into account multiple citation counts. Matsushita, however, moves closer to Toyoda Gosei. Obviously, these two players cite patents from each other frequently, indicating a close technological relationship and hence, competition. In fact, Toyoda Gosei and Matsushita occupy different positions in the LED production value chain: Toyoda Gosei supplies blue LED chips to Matsushita that in a further production step are transformed into white LEDs [23].

*{insert Figure 9 and 10 about here}*

In Figure 3 it can be seen that Sumitomo cooperated with Hitachi, Mitsubishi, NEC, Sharp, and Sony. In Figure 10, all these companies are situated in the lower right "corner" of the network's centre, except Sony which is located more in the upper right half of the network. Hence, Sumitomo seems to have cooperated with a range of companies that possess technologically distinct capabilities.

In general, many companies are connected via citation ties in the network. Figure 11 provides an overview of the extent to which an applicant in the network is cited at least once by other applicants. Nichia, as the most prominent company, is cited by 79% of all other applicants, followed by Toyoda Gosei (77%), Toshiba (70%), and Matsushita (62%). These four applicants are clearly ahead of all others, with Sharp coming next with 49%. This relative citation impact should enhance classical citation analysis that solely counts the absolute number of citations an applicant receives without telling anything about the impact within the competitive environment, namely the citation ties within the technology field in which the applicant is situated.

*{insert Figure 11 about here}*

Figures 9 and 10 are somewhat overloaded with citation ties so that relationships among the most prominent applicants in the centre are difficult to recognize. Figure 12 (a) resolves this issue. Here, only some of the most active applicants from Figure 10 were kept activated in Netdraw. It can be seen now that the strongest inter-applicant citation ties can be found between Toyoda Gosei and Nichia that already were involved in patent litigation [24]. Strong ties also exist between Toyoda Gosei and Matsushita (as explained earlier), and Toshiba and Nichia. Relatively weak ties can be found between Mitsubishi, Sumitomo, and Showa Denko. It therefore seems that these companies are technologically active in separate areas. Relatively weak ties between applicants that are situated relatively close together in the network in Figure 9 and 10 seem, however, somewhat surprising. The reason is that the position of nodes is relative, i.e. determined by all applicants found in the network. Since Figure 12 (a) is determined by the overall sample of applicants in the network, distances are biased when taking only a subset of applicants into account. The bias can be removed by redrawing Figure 12 (a), as was done in Figure 12 (b). The forces in the spring embedding algorithm readjust the distances, and now it becomes clearer that in fact the distance between Nichia and Cree is relatively short, while the distance between Toshiba and Cree is rather large.

Self-citations are also more easily visible in Figures 12 (a) and (b), indicating that Toyoda Gosei seems to possess an outstanding self-citation network. In this context, Figures 9-10, and 12 (a/b) complement the results from Figures 7 and 8.

*{insert Figure 12 about here}*

## 5. Conclusions

The present paper reviews the literature on social network analysis and patent information. Social network analysis is proposed as a tool to improve current visualization techniques in patent analysis that explain cooperation and citation links between inventors, authors, and documents (in this case, patent families). The visualizations presented in the previous section not only include ties and nodes (e.g. inventors or authors), but also additional information such as citation frequency or activity of the nodes, embedding further dimensions into one chart in order to enhance the interpretation of data.

The power of network analysis and visualization techniques is exemplified for the field of III-nitride semiconductor light emitting devices. Cooperation networks between inventors and patent applicants are shown. Due to the large size of the technology field and hence the network, the investigation is conducted only for highly active inventors or applicants. When the activity threshold level is reduced, i.e. the number of patents an inventor or applicant needs to hold in order to be included into the visualization, then subnets of inventors can be identified consisting mainly of colleagues from the same firm.

It is furthermore found that inventors spanning bridges between different inventor groups hold patents that are technologically broader, i.e. possess more IPC classes. Since the data in the analysis is somewhat biased through type-II errors (homonyms, i.e. same name but different persons), this subject deserves more attention in future research.

Citation networks of patent documents highlight not only frequently cited patent families but also those citing other patent families extensively. In our example, the latter case relates to a patent family by Rohm. This patent family is technologically very broad, citing more than 80 patent families within the network. Nichia, who was frequently cited by this patent family, sued Rohm for infringement. Network analysis for applicants that had entered cross-license agreements for white LEDs with Nichia confirmed the key position of the firm and untangled self-citation networks for Toyoda Gosei and Cree, hence, representing company-specific patent thickets.

Even though some applicants possess quite central positions in the applicant citation network, a detailed look reveals that some applicants situated closely together only have weak citation ties, while it is the other way round with others. Strong ties in this context imply either cooperation (Toyoda Gosei and Matsushita) or strong competition (infringement trial between Nichia and Toyoda Gosei). For a technology field, the measure “relative citation impact” is introduced, relating to the number of applicants citing a particular applicant in within a network. Since this measure directly relates to (competing) players, it should much better describe citation impact of a firm and may turn out to be a useful measure in the future.

To conclude, network graphs are not only helpful in human resource management when it comes to assign inventors on R&D teams; they are also helpful when searching for partners in R&D projects, in competitor analysis, due diligence, and many other fields.

The methodologies employed in this paper open up several avenues of future research. Combining both WPINDEX and DPCI inventor networks allows further study of so-called key inventors [1, 25] by extending current bibliometric measures towards indicators resulting from social network analysis, e.g. centrality, betweenness, etc. Some studies already have begun to look into this field [11-12]. Network position, in particular in complex product industries, could furthermore indicate the degree of commercial success of firms. Finally, dynamic approaches can show the emergence of technological trajectories and technological diffusion [8, 26].

References

[1] Sternitzke C, Bartkowski A, Schwanbeck H, Schramm R. Patent and Literature Statistics – the Case of Optoelectronics. World Patent Information 2007;.

[2] Toelle J, Niggemann O. Supporting intrusion detection by graph clustering and graph drawing. Proceedings of Third International Workshop on Recent Advances in Intrusion Detection RAID, October 2000, Toulouse, France. 2000;.

[3] Leydesdorff L, Vaughan L. Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. Journal of the American Society for Information Science 2006;57:1616-1628.

[4] von Wartburg I, Teichert T, Rost K. Inventive progress measured by multi-stage patent citation analysis. Research Policy 2005;34:1591-1607.

[5] Yoon B, Park Y. A text-mining-based patent network: Analytical tool for high-technology trend. Journal of High Technology Management Research 2004;15:37-50.

[6] Calero C, van Leeuwen TN, Tijssen RJW. research cooperation within the bio-pharmaceutical industry: Network analyses of co-publications within and between firms. Scientometrics 2007;71:87-99.

[7] Wagner CS. Six case studies of international collaboration in science. Scientometrics 2005;62:3-26.

[8] Ramlogan R, Mina A, Tampubolon G, Metcalfe J. Networks of knowledge: The distributed nature of medical innovation. Scientometrics 2007;70:459-489.

[9] Clarkson G. Objective Identification of Patent Thickets: A Network Analytic Approach. Harvard Business School Doctoral Thesis, 2004. Available from

http://www.si.umich.edu/stiet/researchseminar/Fall%202004/Patent%20Thickets%20v3.9.pdf.

[10] Fleming L, King Charles III.; Juda, Juda A. Small Worlds and Innovation. Harvard Business School, 2004. Available from http://www.bus.umich.edu/Academics/Departments/CSIB/CSIB/Fleming_04-15-05_Seminar_Paper.pdf.

[11] Salmenkaita J. Intangible capital in industrial research: effects of network position on individual inventive productivity. In: Bettis R (Eds.). Strategy in transition, Malden, Mass. [u.a.]: Blackwell, 2004.

[12] Nerkar A, Paruchuri S. Evolution of R&D Capabilities: The Role of Knowledge Networks Within a Firm. Management Science 2005;51:771-785.

[13] Huisman M, van Duijn MA. Software for social network analysis. In: Carrington PJ, Scott J, Wasserman S (Eds.). Models and Methods in Social Network Analysis, Cambridge University Press, 2005.

[14] Batagelj V, Mrvar A. Pajek - Program for large network analysis (as of Mar 12, 2007). Available from http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

[15] Borgatti SP, Everett MG, Freeman L. Ucinet 6 for Windows - Software for Social Network Analysis. Harvard, MA: Analytic Technologies 1999;.

[16] Golbeck J, Mutton P. Spring-embedded graphs for semantic visualization. In: Geroimenko V, Chen C (Eds.). Visualizing the semantic web - XML-based internet and information visualization, London: Springer, 2006.

[17] Lanjouw JO, Schankerman M. Symposium on the patent system and innovation - Characteristics of Patent Litigation: A Window on Competition. Rand Journal of Economics 2001;32:129-151.

[18] Harhoff D, Reitzig M. Determinants of opposition against EPO patent grants-the case of biotechnology and pharmaceuticals. International Journal of Industrial Organization 2004;22:443-480.

[19] Lerner J. The importance of patent scope: an empirical analysis. Rand Journal of Economics 1994;25:319-333.

[20] Freeman LC. Centrality in social networks: conceptual clarification. Social Networks 1979;1:215-239.

[21] Lumileds. Corporate Website (as of Mar 19, 2007). Available from http://www.lumileds.com.

[22] LEDs Magazine. White LED patents (as of Mar 19, 2007). Available from http://www.ledsmagazine.com/news/4/1/32/WhiteLEDpatents.

[23] Hara Y. White LED lamp market brightens (as of 18-Jul-2002). Available from http://www.eet.com/pw/fo/OEG20020718S0013.

[24] Japan Corporate News Network. Toyoda Gosei Announces Overall Settlement Concerning Blue LED Lawsuits (as of 17-Sep-2002). Available from http://www.japancorp.net/Article.Asp?Art_ID=3852.

[25] Ernst H. Key Inventors: Implications for human resource management in R&D. In: Kocaoglu DF, Anderson TR (Eds.). PICMET '99 Portland International Conference on the Management of Engineering and Technology, Portland: 1999.

[26] Mina A, Ramlogan R, Tampubolon G, Metcalfe JS. Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge. Research Policy 2007;in press:doi:10.10.1016/j.respol.2006.12.007.

Figure 1: Inventor network. Source: Patent families from WPINDEX. Graphed with spring embedding algorithm. Circle size: number of patent families.

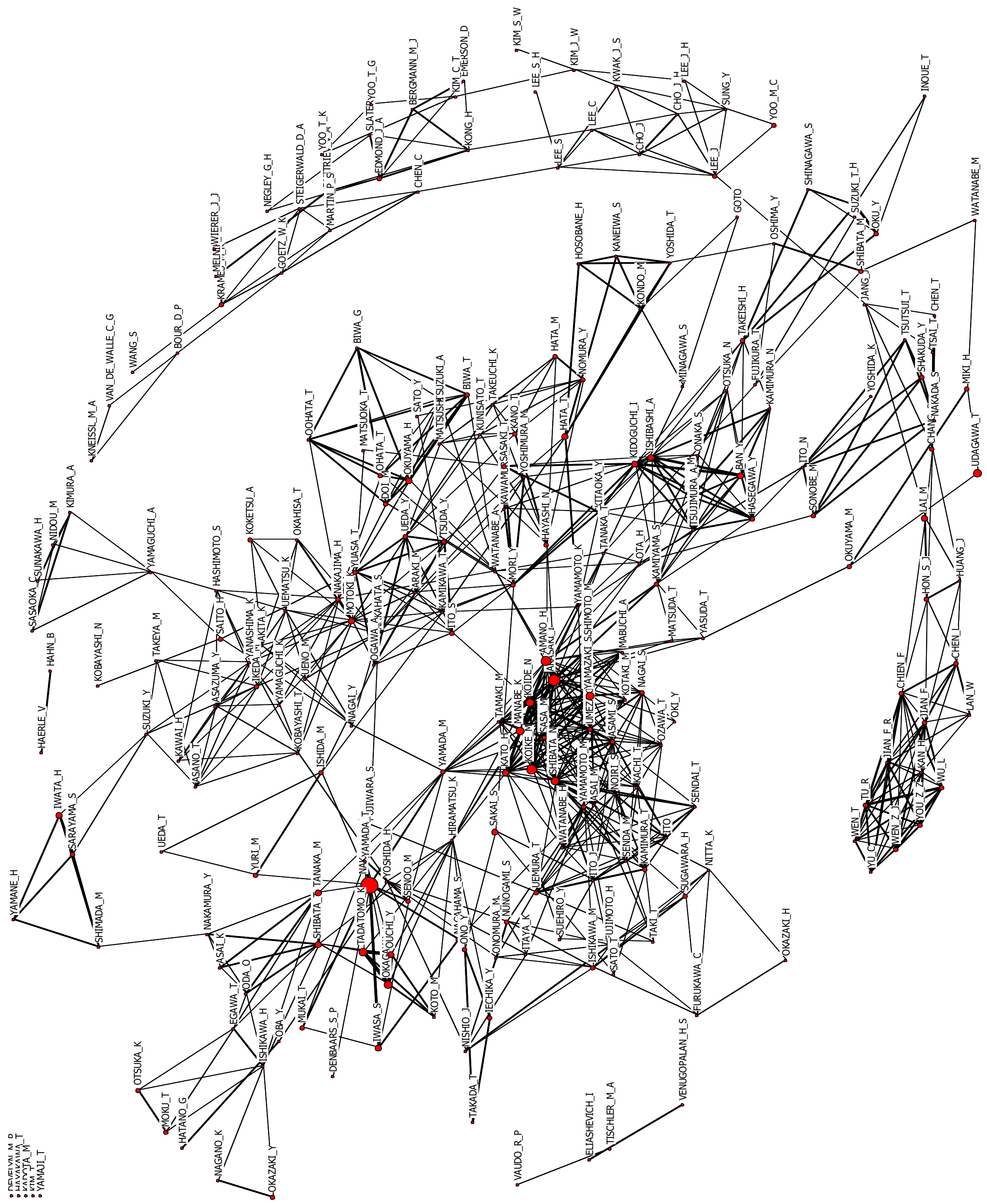


Figure 2: Egonets of Nakamura (a) and Akasaki (b) (excerpt from Figure 1).

(a) (b)

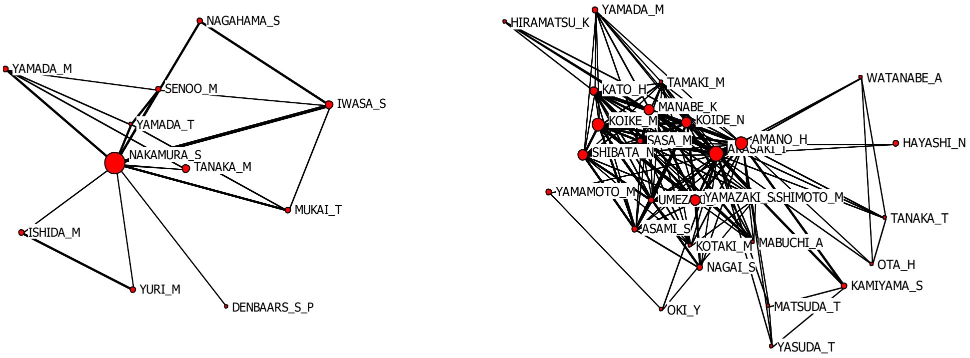



Table 1: Pearson correlation coefficient between normalized centrality betweenness and the number of different IPC classes per patent family and inventor. N=240.

| (1) 4-digit IPC class | (1) | (2) | (3) |
|---|---|---|---|
| (2) 7-digit IPC class | 0.843** | | |
| (3) full IPC class | 0.793** | 0.842** | |
| (4) centrality betweenness | 0.025 | 0.006 | 0.142* |

** significant at the 1% level; * significant at the 5% level

Figure 3: Applicant network. Source: Patent families from WPINDEX. Graphed with spring embedding algorithm. Circle size: number of patent families.

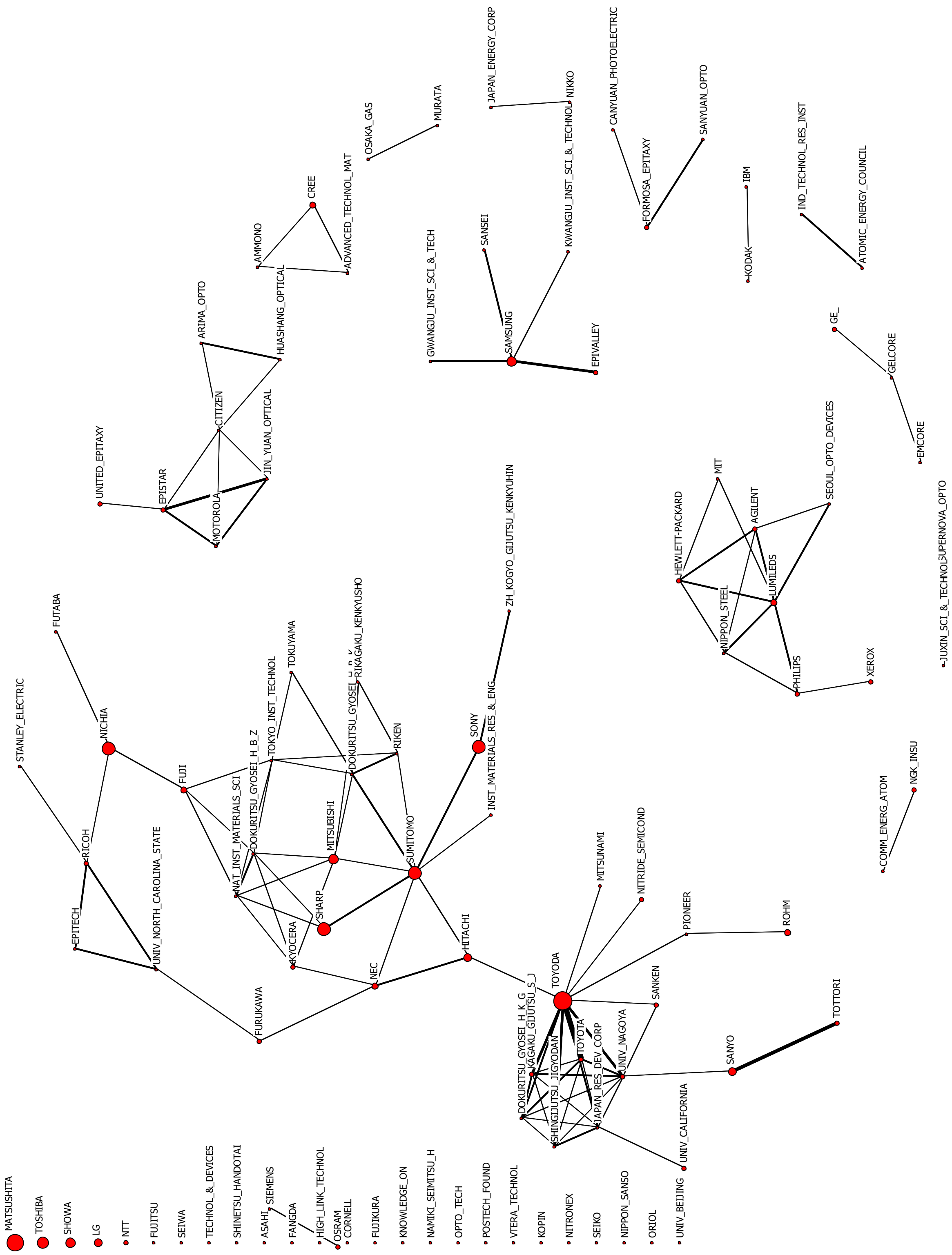


Figure 4: Inventor network. Source: Patent families from DPCI. Graphed with spring embedding algorithm. Circle size: average citations received per patent family.

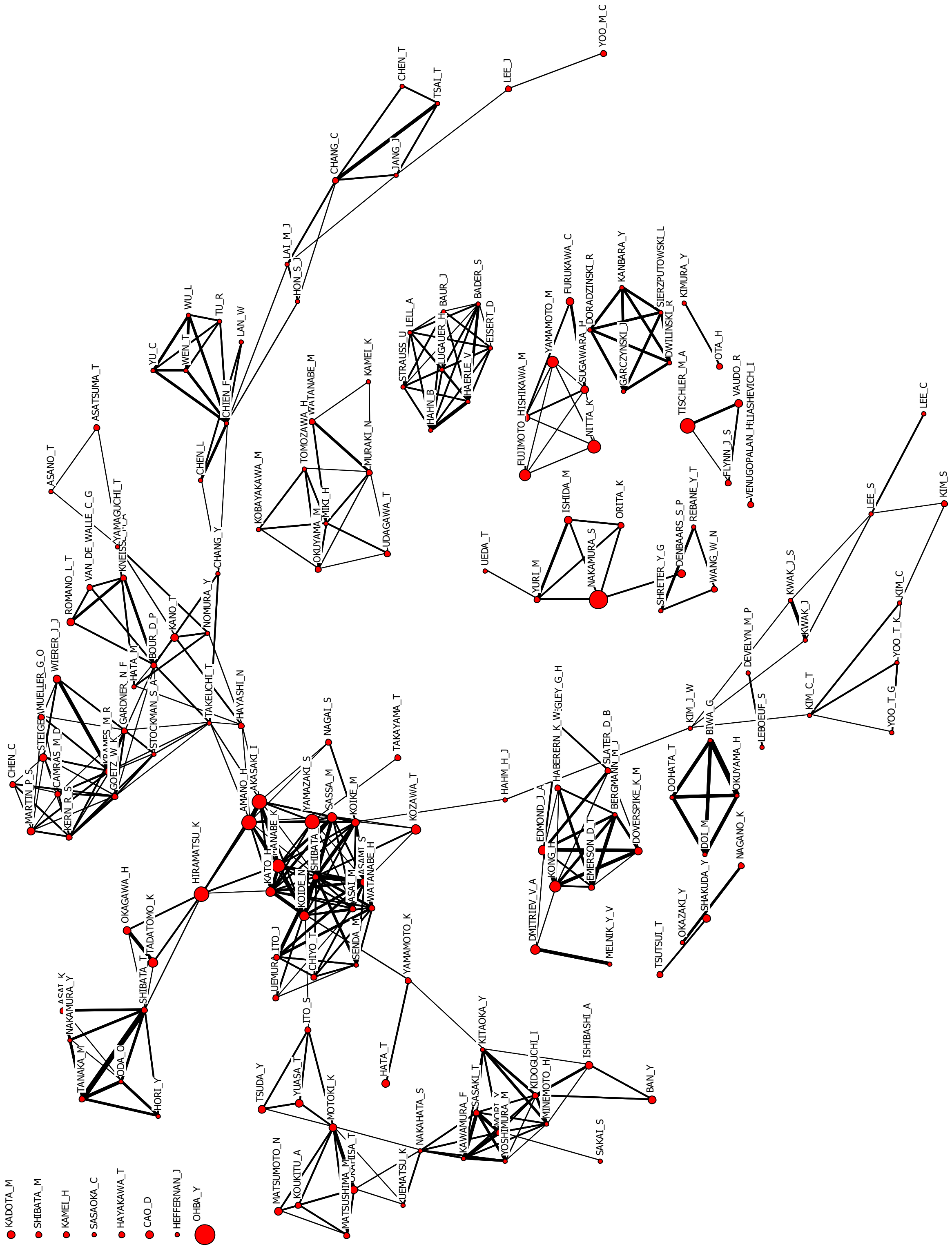


Figure 5: Patent citation network. Source: Patent families from DPCI. Timeframe: 1990-1995. Graphed with spring embedding algorithm. Node size: patent age (large = old).

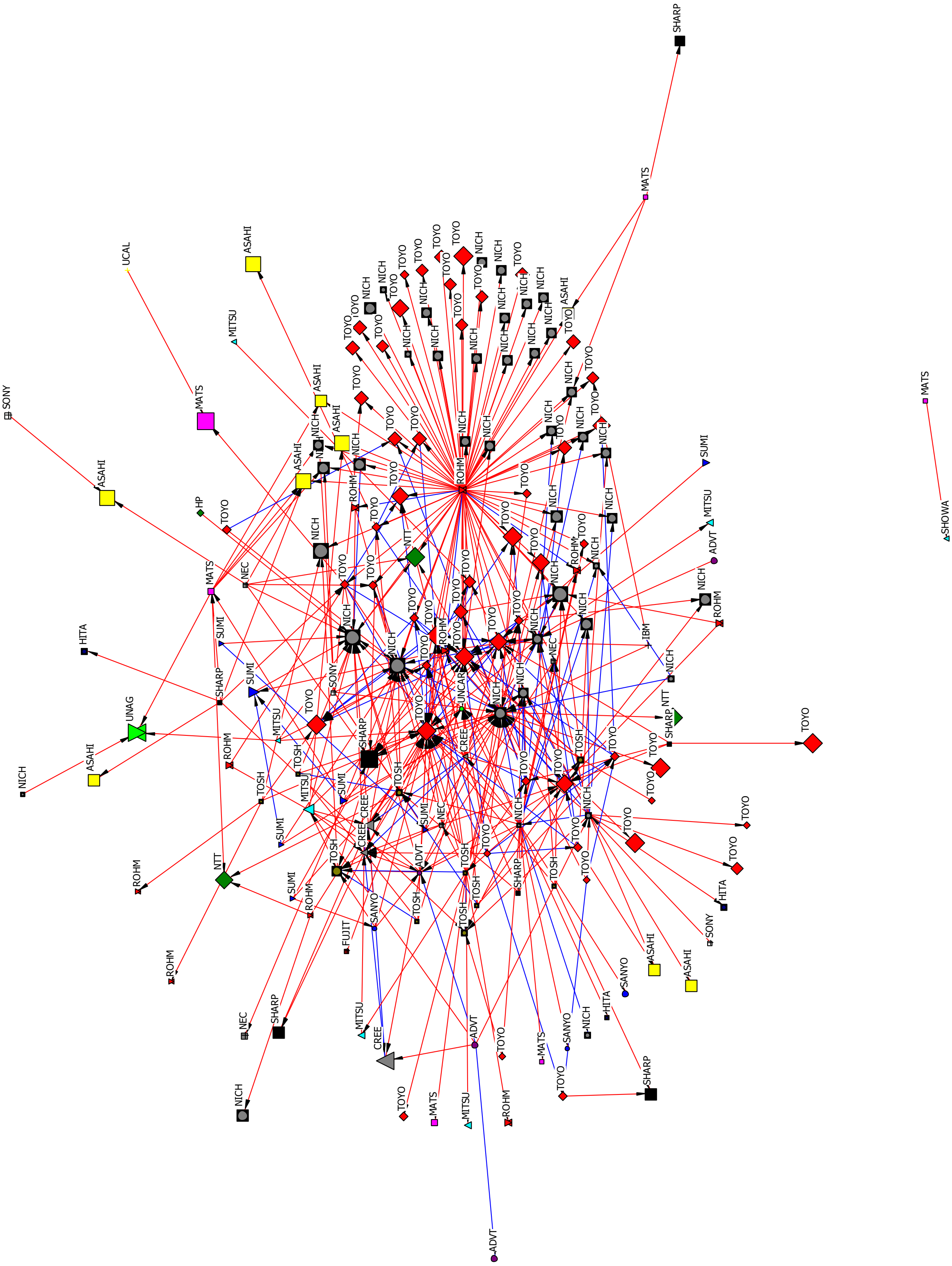


Figure 6: Ego-patent-citation-network for Rohm patent. WPINDEX AN 1996-244765 (e.g. JP08097470) (excerpt from figure 5).

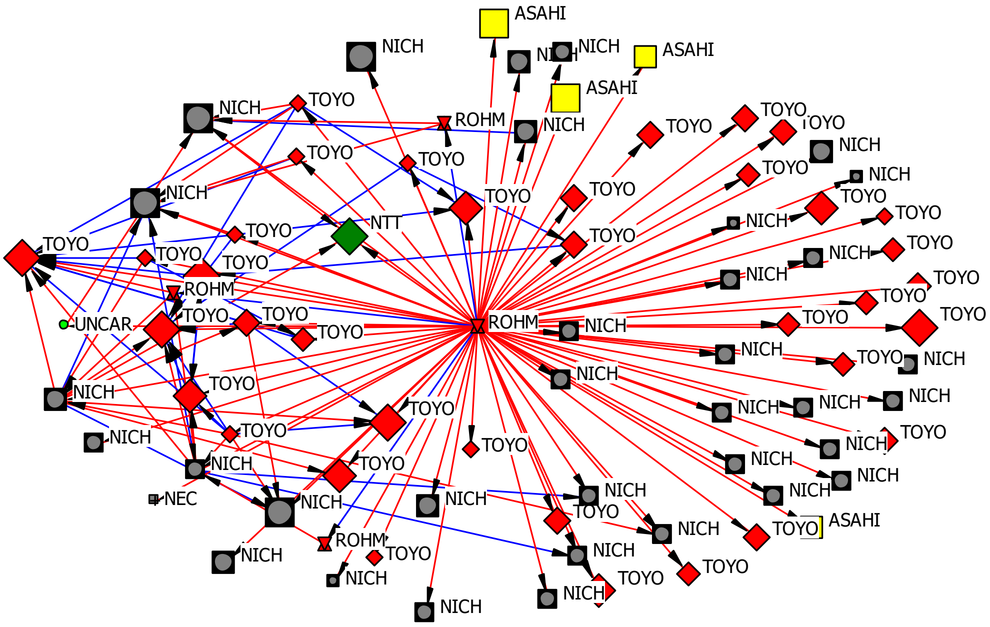

ASAHI
NICH
TOYO
ROHM
NTT
UNCAR
NEC

Figure 7: Patent citation network for Cree, Lumileds, Nichia, Osram, and Toyoda Gosei. Source: Patent families from DPCI. Graphed with spring embedding algorithm. Node size: Citation frequency. Line color: black = self citations, grey = foreign citations.

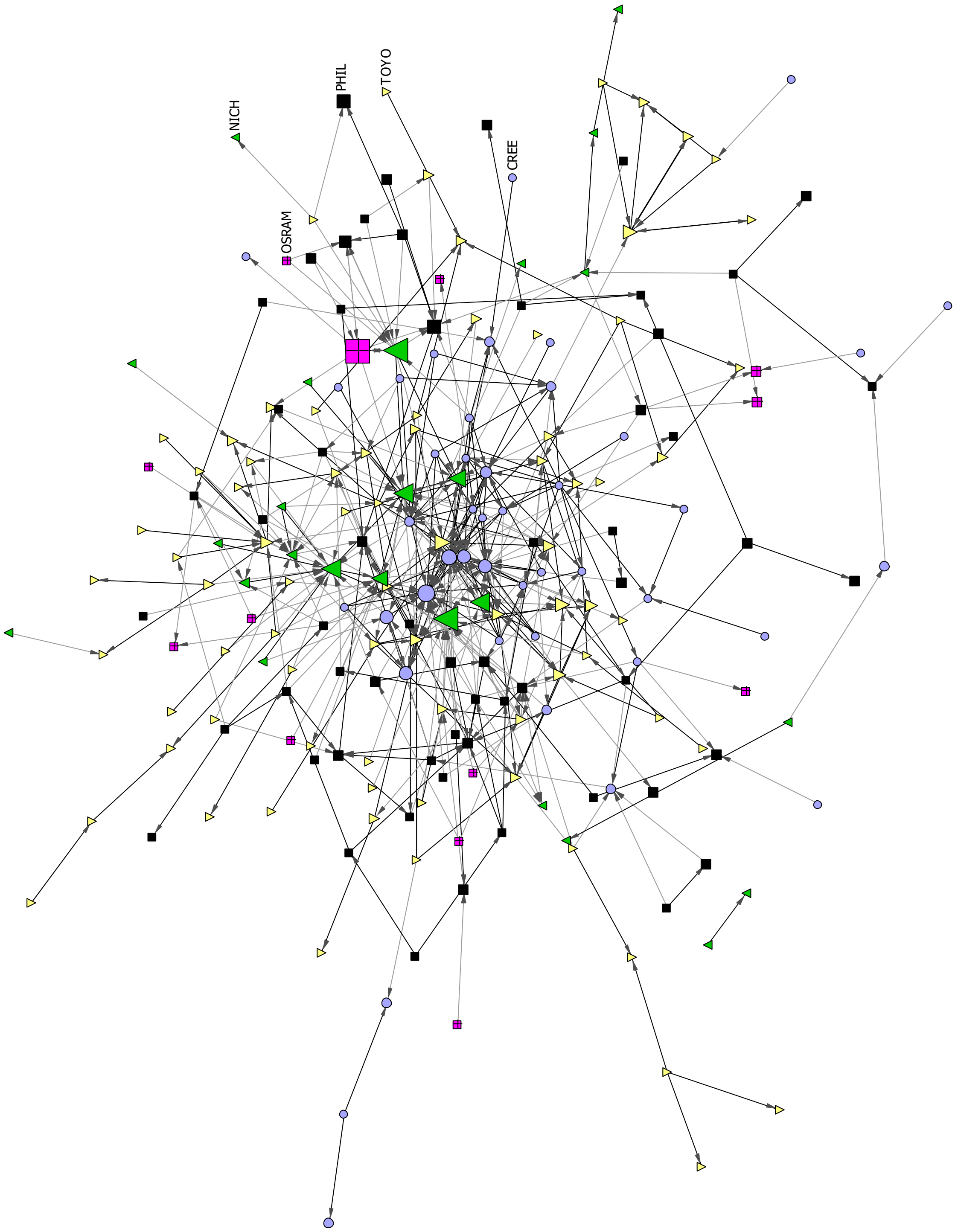


Figure 8: Patent citation network for Cree. Source: Patent families from DPCI. Graphed with spring embedding algorithm. Circle size: absolute citation frequency.

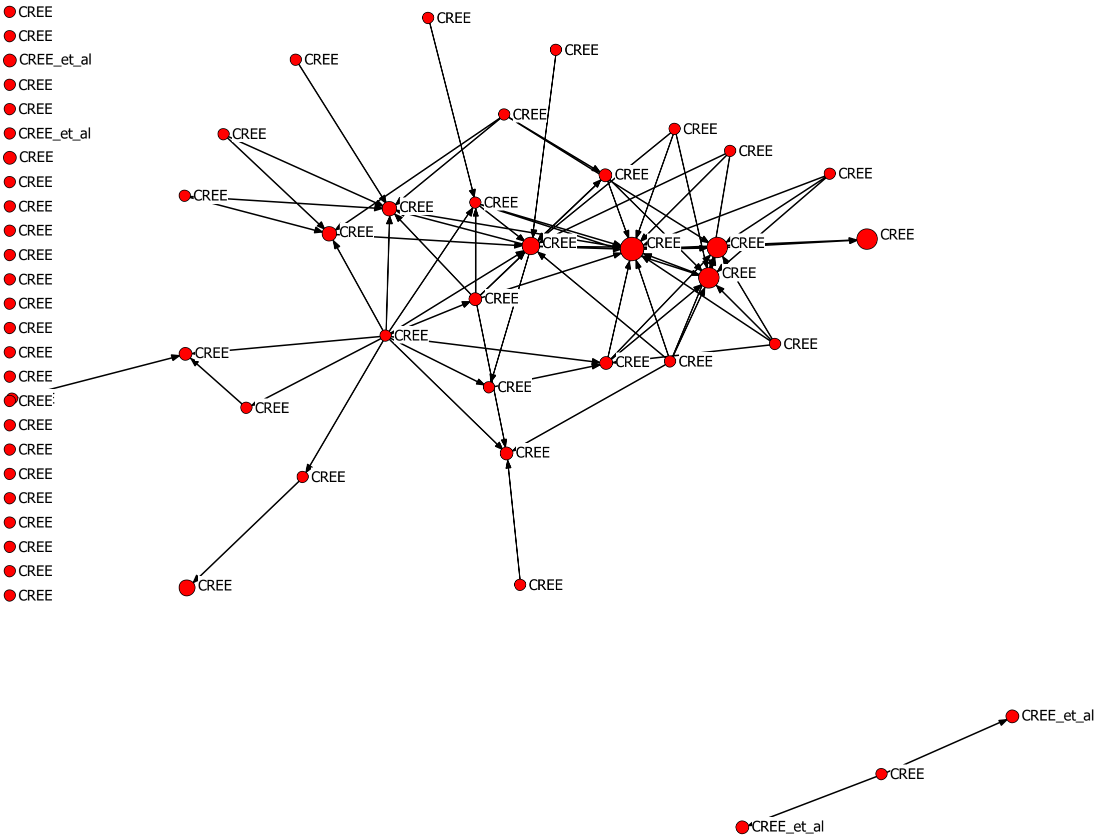


Figure 9: Patent applicant citation network. Only single citations of documents by applicant. Source: Patent family size from DPCI. Graphed with spring embedding algorithm. Circle size: patenting activity.

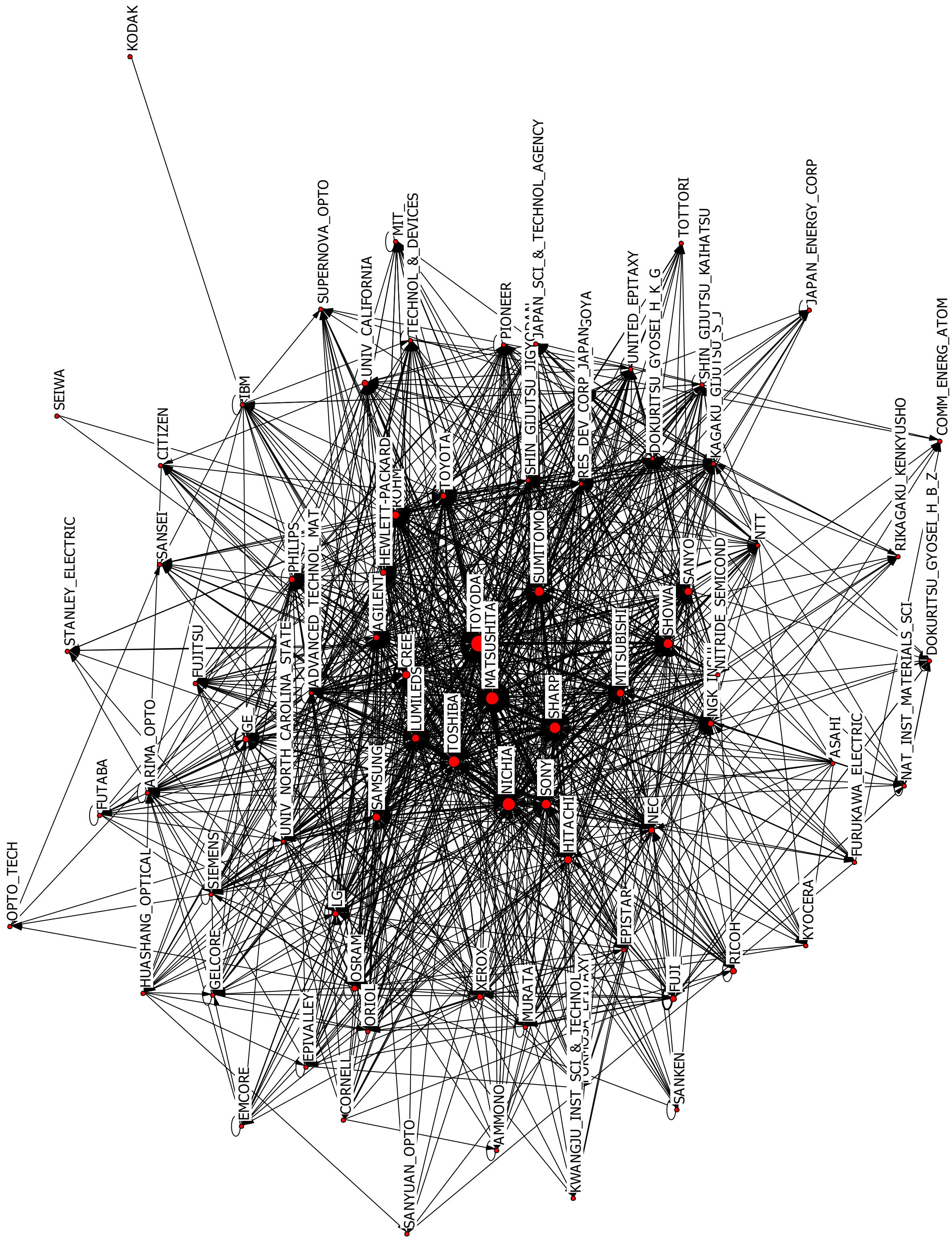


Figure 10: Patent applicant citation network. Multiple citations of documents by applicant possible. Source: Patent families from DPCI. Graphed with spring embedding algorithm. Circle size: patenting activity.

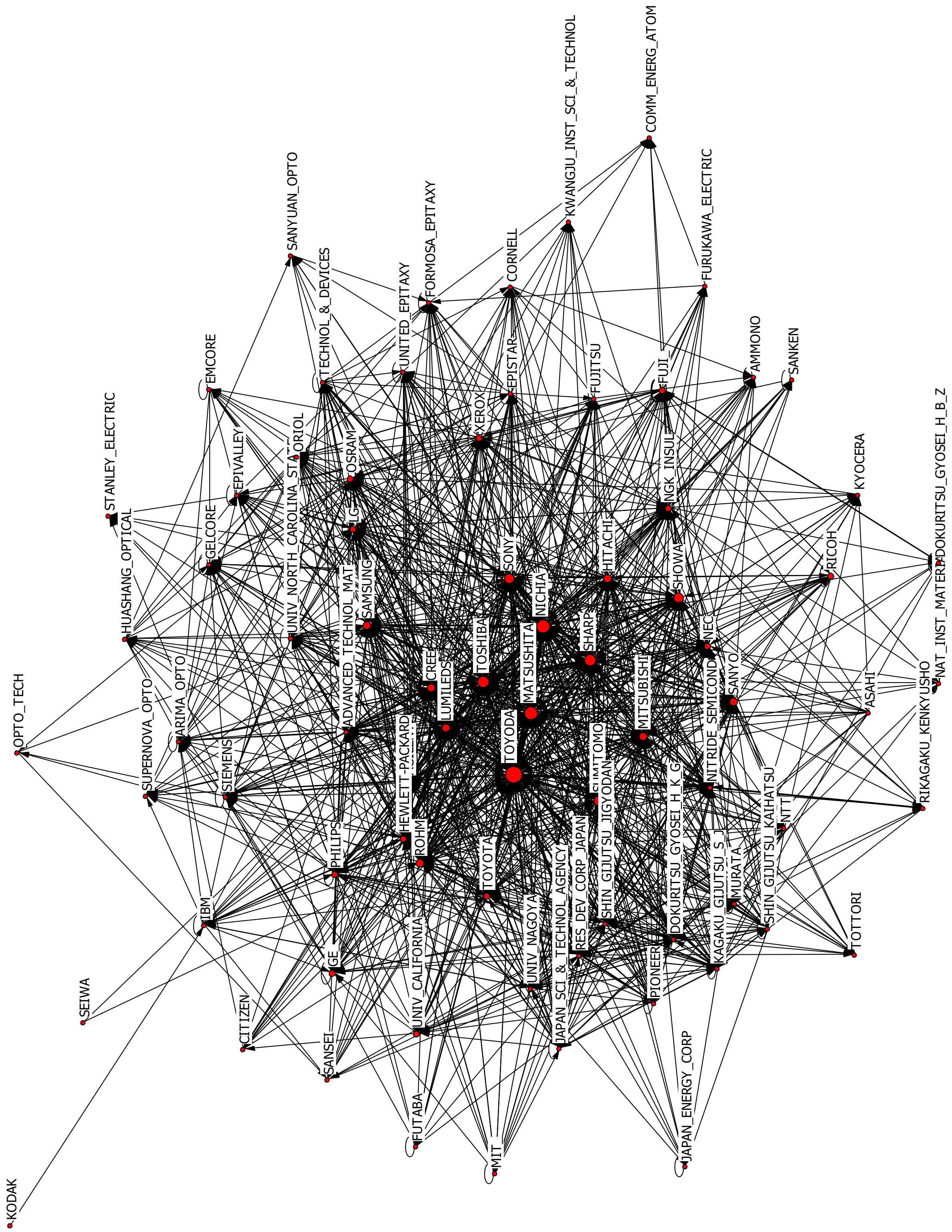


Figure 11: Ranking of applicants regarding their citedness by other applicants within the network.

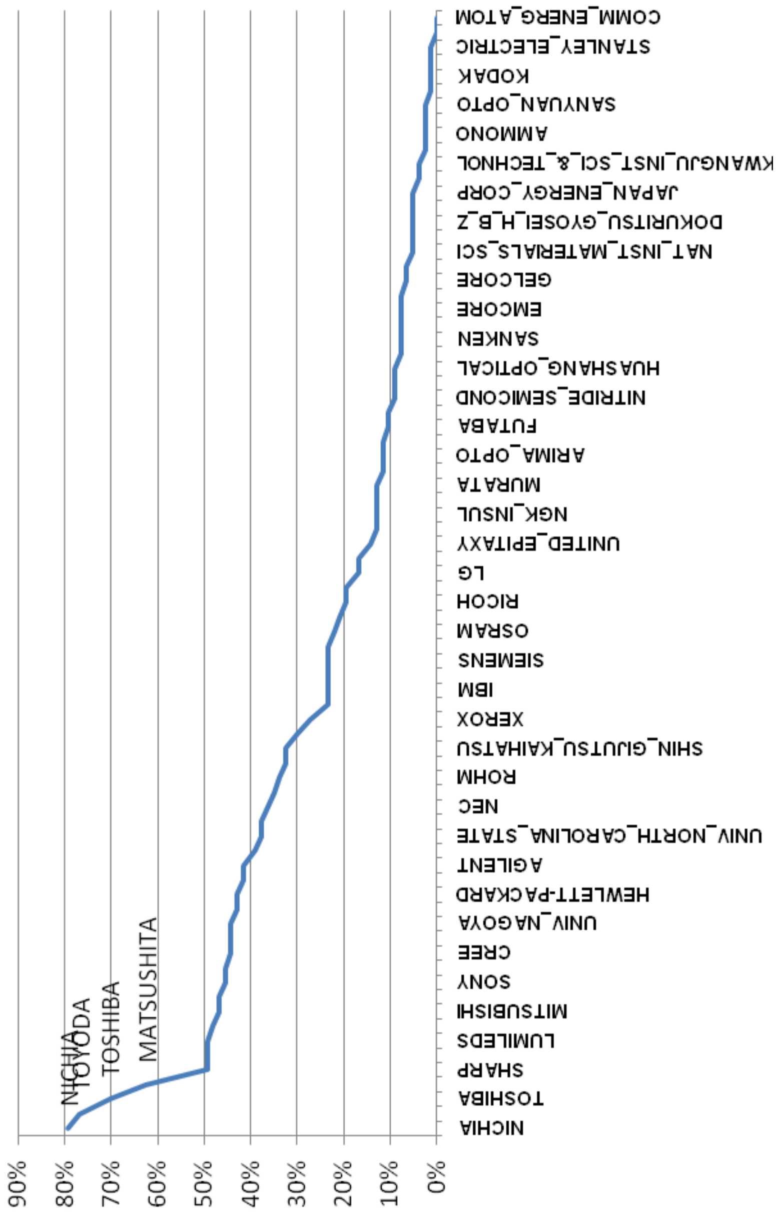
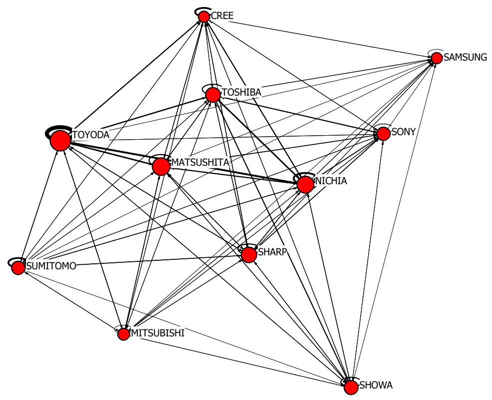
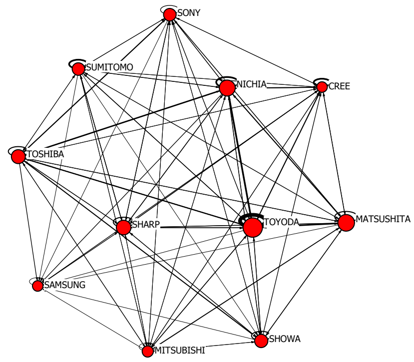


Figure 12: Patent applicant citation network. (a) Excerpt from figure 9. (b) Chart in (a) redrawn. Source: Patent families from DPCI. Graphed with spring embedding algorithm. Circle size: patenting activity.

(a)




(b)



## Appendix

Table A1: Abbreviations of applicant names shown in Figures 5-7.

| Abbreviation | Name | Abbreviation | Name |
|---|---|---|---|
| TOYO | TOYODA | FEPI | FORMOSA EPITAXY |
| MATS | MATSUSHITA | UNAG | UNIV NAGOYA |
| NICH | NICHIA | EMCO | EMCORE |
| SHARP | SHARP | GELC | GELCORE |
| TOSH | TOSHIBA | FUTA | FUTABA |
| SHOWA | SHOWA | SHIN1 | SHIN GIJUTSU JIGYODAN |
| SONY | SONY | ASAHI | ASAHI |
| SUMI | SUMITOMO | KYOC | KYOCERA |
| LUMI | LUMILEDS | MURA | MURATA |
| SANYO | SANYO | TDEV | TECHNOL & DEVICES |
| MITSU | MITSUBISHI | ADVT | ADVANCED TECHNOL MAT |
| CREE | CREE | KAGA | KAGAKU GIJUTSU S J |
| SAMS | SAMSUNG | NATI | NAT INST MATERIALS SCI |
| HITA | HITACHI | EPIS | EPISTAR |
| ROHM | ROHM | SUPN | SUPERNOVA OPTO |
| NEC | NEC | MIT | MIT |
| FUJI | FUJI | SEIWA | SEIWA |
| OSRAM | OSRAM | SIEM | SIEMENS |
| TOYT | TOYOTA | JAPA | JAPAN SCI & TECHNOL AGENCY |
| XEROX | XEROX | TOTT | TOTTORI |
| LG | LG ELECTROICS | DOKU | DOKURITSU GYOSEI H B Z |
| NGK | NGK INSUL | KODAK | KODAK |
| RICOH | RICOH | RESJ | RES DEV CORP JAPAN |
| AGIL | AGILENT | UNCAR | UNIV NORTH CAROLINA STATE |
| HP | HEWLETT-PACKARD | AMMO | AMMONO |
| PHIL | PHILIPS | CORN | CORNELL |
| GE | GENERAL ELECTRIC | ORIOL | ORIOL |
| UCAL | UNIV CALIFORNIA | RIKA | RIKAGAKU KENKYUSHO |
| EPIV | EPIVALLEY | SANS | SANSEI |
| SANK | SANKEN | STAN | STANLEY ELECTRIC |
| UEPI | UNITED EPITAXY | CITI | CITIZEN |
| NISEM | NITRIDE SEMICOND | HUAS | HUASHANG OPTICAL |
| PION | PIONEER | SHIN2 | SHIN GIJUTSU KAIHATSU |
| FURU | FURUKAWA ELECTRIC | COMM | COMM ENERG ATOM |
| NTT | NTT | JAPC | JAPAN ENERGY CORP |
| DOKU-G | DOKURITSU GYOSEI H K G | KWAN | KWANGJU INST SCI & TECHNOL |
| ARIMA | ARIMA OPTO | OPTO | OPTO TECH |
| FUJIT | FUJITSU | SANYU | SANYUAN OPTO |
| IBM | IBM | *el_al | Patent resulting from a cooperation. The most active applicant is named first. |