

# Scientific Progress as Diffusion on a Social Network: an Empirical Case Study

*Although the spread of scientific theories is a prerequisite for scientific progress, little is known about the way in which science spreads within the scientific community. To understand and stimulate scientific progress it is therefore key to understand how a scientific hypothesis diffuses in the scientific community. In this paper, I analyze the diffusion of one particular scientific hypothesis, namely the hypothesis on the strength of weak ties, introduced by Granovetter in 1973. From data on 6296 scientists and their reference behaviour, a network is constructed that reveals the diffusion process. It is found that the hypothesis is used by different communities of scientists, each with their own key-narrative into which the hypothesis is fit. The diffusion within the communities shares similarities with the diffusion of an innovation in which innovators, early adapters and a majority can be distinguished. Furthermore, it is found each community contains a hub scientist that is disproportionately often referenced and is largely responsible for carrying the hypothesis into his scientific subfield. As a result, each community can be represented by its hub scientists. These results suggest a clear diffusion pattern that can be studied further. Not only can these insight be used to encourage scientific progress, they also give valuable input for the study of diffusion in general.*

## 1. Introduction

Our society is influenced by scientific progress in so many ways (Phelps et al, 2012), yet it remains largely unknown how this progress takes place. In fact, scholars even disagree about the definition of scientific progress. A point of agreement, nevertheless, is that the spread of scientific theories is a prerequisite for progress. Just like technological innovations, an invention does not imply an innovation. Innovations refer to the use of novel ideas or methods and not just to the creation of them. Diffusion is hence a necessary ingredient. To understand and stimulate scientific progress it is therefore key to understand how a scientific hypothesis diffuses in the scientific community.

Diffusion, in turn, is a very broad term, being studied by scientists of various research fields. If the diffusion of science were to 'fit' into one specific diffusion research field, it is not immediately clear where it belongs. Instead, it is more likely that we can use results from several research fields in order to understand diffusion in science. Similarly, explicitly studying the diffusion of a scientific hypothesis can give relevant input for multiple research fields.

Below, findings of the three scientific fields on which this research builds and to which this research is most relevant will be discussed, namely diffusion of innovations, narratives in science and the citation structure of the scientific community.

First, this research falls into the field of diffusion of innovations, initiated in the 1960s by publications of Fourt and Woodlock (1960), Manfies (1960), Floyd (1962), Rogers (1962) Chow (1967) and Bass (1969), (Maede & Islam, 2006). Diffusion of innovation theory originally focused on technological innovations, but has since been applied to many other fields, amongst others, scientific growth (Valente & Rogers, 1995). Characteristic of the diffusion of innovation theory is the logarithmic function as a growth curve and the corresponding classification of innovators, early adapters, majority and laggards. Although these different groups are later named differently, the main function remains; innovators are the ones to come first with a novel idea or method, but the idea needs to be picked up by some early adapters, otherwise known as opinion leaders or influencers, to be adapted later on by the majority. This theory forms the first part of the hypothesis of this paper, stating that for a scientific hypothesis to diffuse successfully, it needs to be adopted by a prominent scientist, an opinion leader.

The second finding on which this research builds concerns narratives. For a literature review in, again, the field of diffusion of innovations Greenhalg and colleagues (2005) mapped all research activities within this extremely large field. They took as the main unit of analysis the unfolding 'storyline' of a research tradition over time (2005) and identified 13 different such key-narratives from the literature. Researchers in these different narratives had conceptualised, explained and investigated diffusion of innovations in different ways. Moreover, Greenhalg and colleagues find that each narrative contains some initial 'breakthrough' paper that attracts promising young scientists to the new field. Generalizing this finding, this research expects that a scientific theory that is widely spread is

always translated to fit the key-narratives of a scientific field.

Last, quantitative research on the structure of science finds that science in general shows a very uneven distribution of references (de Solla Price, 1976). Some scientists and scientific papers are referenced disproportionately often, whereas the large majority is cited only once or twice. These uneven distributions, specifically the power law, are shown to be present within the different scientific communities as well. In a collaboration network based on co-authorship Neman (2004) shows that within the scientific fields considered in his research, including statistical physics and mathematical ecology, each field forms their own scientific community and within each community the number of collaborators per scientist follows a power law.

Mapping these three results together, three laws emerge for the diffusion process of a scientific hypothesis. First, for a scientific hypothesis to spread into a new research field, it needs to be interpreted and conceptualised to fit into the larger narrative of the new field. Once this narrative is set successfully, by or picked up by some prominent scientists, the hypothesis can spread. Due to the reference behaviour of scientists, clusters are formed around the prominent scientist. Finally, these highly referenced scientists can then be seen as the representatives of the clusters.

This paper gives the results of a case study to explore the diffusion of science and investigate to what extent it is governed by these three laws. In this case study the diffusion of a particular scientific hypothesis is tracked, namely the hypothesis on the strength of weak ties, introduced by Granovetter in 1973. Granovetter's paper has been referenced over 6.900<sup>1</sup> times by scientists of more than 81 different scientific fields<sup>2</sup>. From neurobiology to literature studies, scientists have found his hypothesis useful and applicable in their field. This wide spread makes the hypothesis a particular good case study for the diffusion of science.

In short, Granovetter's hypothesis states that the people forming a bridge between two different communities in a social network are the ones that enable information to spread widely. These people are often weak ties<sup>3</sup>, like acquaintances, as opposed to strong ties, like friends. Hence the strength of weak ties.

The approach of this research is as follows; from all the scientists referencing to 'The Strength of Weak Ties' in the period 1973-2012 a network is created that develops over time. With every new scientist that enters the network, edges are drawn from this new scientist to the earlier scientists who the new scientist references in his first paper referencing Granovetter. The resulting network thus visualises all scientific activity that uses the hypothesis on the strength of weak ties. The evolution of the network represent the diffusion process.

<sup>1</sup>6,900 is a low estimate, based on the articles present in the Web of Science core Collection. Google scholar gives a figure of 28,465.

<sup>2</sup>This is based on the scientific field – categorization made by Web of Science. Following other scientific databases, such as Scopus, lead to a similar number.

<sup>3</sup>A weak tie indicates a relation between two people in which the two people do not see each other very often, and are not very similar in terms of their behavior, interests and friends. Especially this last characteristic is important in Granovetter's hypothesis which relies on the fact that if A has a strong tie with B and with C, more often there is also a tie, strong or weak, connecting B and C.

This approach has, to my knowledge, never been taken before. This research should therefore be seen as an experimental first try for a research technique in diffusion on social networks, in particular a scientific network, that holds potential for the future. An interesting next step would be to do a similar analysis as is done here with the diffusion of another scientific hypothesis and compare for similarities and differences.

The structure of this paper is as follows. In the next section, I will describe the data and analysis methods used in more detail. Thereafter, the findings of this case study are discussed in two sections. The first section, Section 3, is about the static network in 2012 which visualises all scientific activity that uses the hypothesis on strength of weak ties. The focus is here on community structure in this network and the distribution of the in-degree. The next Section, Section 4, evaluates the evolution of the network in time. The analysis is separated into the development as the network as a whole and the development within communities in the network. The findings described in Section 3 and 4 show that some researchers play a key role in the structure and evolution of the network. Section 5 subsequently gives a close-up of these scientists. Section 6 summarizes the findings of the diffusion of this specific hypothesis for diffusion of scientific hypotheses in general.

## 2. Data & Methods

Data is gathered from the Web of Science core collection. Web of science is the largest online database with scientific articles from the year 1900 till now. The database contains '*The Strength of Weak Ties*' and gives the option to list all articles that referenced this paper. From all these articles and their authors in the period 1973-2012, the following data is gathered:

Article data:

- Title, accession number, journal, publication date, research areas, web of science categories, references

Scientist data:

- Initials, unique ID, articles written

To every scientist I assigned a unique year, namely the year of publication of his first article referencing to Granovetter's 1973 paper. In that particular first paper, the scientist, call him 'Peter' for the moment, also references other papers. Some of these papers reference Granovetter's 1973 paper as well. These papers have hence served as examples of the use of Granovetter's hypothesis for Peter. Therefore, I linked Peter to all scientists that were authors of these example papers.

With this data, a directed network is constructed. The nodes represent the scientists referencing '*The Strength of Weak Ties*'. The edges represent a source relationship: an edge from scientist A to scientist B means that in scientist A's first paper referencing Granovetter, scientist A also referenced a paper by scientist B in which Granovetter's hypothesis is used.

In 2012, the network of scientists referencing '*The Strength of Weak Ties*' consists of 8019 nodes. Note that this figure represents the different scientists, which is higher than the number of articles (6900) mentioned in the introduction. Since the interest of this research is the diffusion of Granovetter's hypothesis all nodes without ties to the giant component are not further considered, leaving the network with 6296 scientists that all play some role or another in the diffusion process.

In addition, every author is classified into one or two research area's based on his first paper referencing Granovetter. Overall in 2012, the largest research areas are business & economics (24.8%), sociology (12.4%), psychology (5.6%) and computer science (5.7%). All other 78 research areas contain less than 5%, see Figure 1 for more details.

Because this network is only based on references around Granovetter's hypothesis, common concepts in networks analysis such as 'shortest path' and 'centrality' are not easily interpreted. To

detect the community structure in the network, discussed in Section 3, the modularity is therefore optimized using the Louvain method of Vincent D Blondel & Jean-Loup Guillaume's (Blondel et al, 2008). This algorithm does not make use of any concepts with ambiguous interpretation and therefore seems suitable.

## 3. Network structure in 2012

The first analysis of the data is on the static network in 2012. I focus on two main characteristics of the network in 2012 that are relevant for the hypotheses of this paper and have a clear interpretation, namely the community structure and the distribution of the in-degree.

### Community Structure

From the narrative perspective, a scientific theory does not stand by itself. The theory always fits into a larger story. Greenhalg and colleagues (2005) showed that with the widely spread theory of diffusions of innovations, there exists multiple key-narratives in which the theory is conceptualised, explained and investigated differently.

Granovetter's hypothesis on strength of weak ties has also diffused to various research fields. The hypothesis of this paper is that these different fields hold their own debates and Granovetter's hypothesis fits into the key narratives in different manners. Scientists of different debates hence have little to share with one another.

In this network an edge between scientist A and B implies that scientist B has read and used scientist A's paper in his first paper referencing Granovetter. If there are indeed these different debates in different scientific fields, it is expected that the network is clustered. Clustering means that the nodes of the network can be grouped into larger community-classes, where the number of edges between nodes of one community is large and the number of edges between nodes of different communities is relatively small. Each community would represent a different scientific narrative, which is strongly correlated with research area. The data of this research also gives the research areas of each scientist, hence I investigate whether there is clustering and whether the clustering correlates with research area.

The data shows that this network can be divided into 17 community-classes. The corresponding modularity has a value of 0.54. Four of these classes contain less than a half percent of all nodes, which means that they consist of less than 31 scientists. I have judged this to be a too small sample to draw any conclusions on and these classes are therefore not considered for further analysis.

Figure 2 shows the network in 2012 with distinct colours for the 13 communities under consideration. Mapping the distribution of research areas within each class shows that the communities are strongly correlated to research field. The red community, class 1, is the largest community containing 18.6% of all scientists. The distribution of research areas in this class, see Figure 3, shows considerably more sociologists and psychologists and less business & economists than the overall network and this deviation proves to be significant<sup>4</sup>. The second largest community is the bright orange class 11, see Figure 4, and consist of considerably more business & economists (total 52.6%) and less sociologists, a significant deviation as well<sup>5</sup>. Likewise, the other communities can be characterized by their research areas.

Ultimately, I find four social classes, red(1)/bright green(16)/purple(9)/yellow-green(12), neighbouring each other and five business & economics classes, blue (5) /light blue (2) /bright orange(11) /dull orange (15) /dark green (8), also neighbouring each other. Located between the boundaries of several communities we find scientists of class 4 and 13, yellow in Figure 2. These two classes are characterized by having many computer scientists and information scientists. There is a large (8.6%) clustered class that is less connected to the rest of the network, class 7, black in Figure 2, which consists of

<sup>4</sup> The Chi-squared value is 587 With 27 degrees of freedom.

<sup>5</sup> The Chi-squared value is 816 with 22 degrees of freedom.

mainly physicists (28.5%), computer scientists (16.2%) and researchers in the field of science & technology (7.9%). Two other of such clustered and less connected classes are class 0, in the upper right corner with old pink colour in Figure 2 and class 10, in the left down corner with white colour in Figure 2. Both classes are specialized; class 0 consists for 27.6% of geographers and contains no sociologists; class 10 holds almost all government & law researchers (33% as opposed to 2% in the overall network) and a large percentage of communication scientists (15.7 %).

In sum, the data shows that there is clustering and the clustering correlates with research area. Two research fields that I broadly categorized as business & economics and social sciences are made up of several classes that cannot easily be distinguished based on their distribution of research areas. I expect that although these classes consists of researchers of approximately the same fields, they differ in the debate to which Granovetter's hypothesis is applied. I make an attempt to show this in section 5.

### In-degree distribution

It is shown in citation networks that citations are not distributed evenly. In fact, it has been shown that some scientists are referenced disproportionately often whereas the majority is only rarely referenced (reference). More formally, the distribution is shown to follow a power law, which is a specific kind of such an uneven distribution. This phenomenon can be explained by the 'cumulative advantage' (de Solla Price, 1976), more generally known as preferential attachment (Barabási, 1999), which in this case refers to the fact that 'popular' scientists that are often cited have a higher probability to be cited again than less known scientists.

I expect to see the same phenomenon in the network based on Granovetter's hypothesis. Note, however, that this network does not show all citations, but only the citations of scientists' first article referencing Granovetter.

In the previous section, we have seen that the network consists of several clusters, representing different scientific fields. In a study of Neman (2004) that also divided a network into communities with different scientific fields, it was shown that in each community the number of collaborators per scientist follows a power law. The same phenomenon can therefore be expected in this network, where the power law, or other uneven distribution, is expected to be found for the in-degree. If this hypothesis holds, each of the found community-classes has an uneven distribution of the in-degree, similar to a power law.

An alternative hypothesis would be the following. Because Granovetter was not well-known before 1973 we can expect that the first to reference him were closely related scientists of the same research field. Therefore, most of the high in-degree scientists are expected to be part of the same class. Moreover, the earlier a scientist references Granovetter, the more time thus probability the scientist has of being referenced again. Scientists that reference Granovetter earlier hence have a larger in-degree. Following this hypothesis, the scientists with the highest in-degree are expected to be connected to earlier years and part of the same research class.

The data of the in-degree shows to be consistent with the expected power law, with an exponent of value 1.16.<sup>6</sup> This fact represents that most scientist are rarely referenced by new scientists, and they can therefore not explicitly be called meaningful for the spread of Granovetter's hypothesis. A small number of scientists, however, is disproportionally often referenced by new scientists entering the network and they are therefore largely responsible for spreading Granovetter's hypothesis.

Selecting the most responsible scientists, the 63 (1%) scientists with the highest in-degree, it is found that these scientists are neither all from early years, nor part of the same class. Figure 5 shows a timeline with corresponding percentage of these high in-degree scientists which clearly illustrates that the scientists most responsible for the spread of Granovetter's hypothesis were not

simply the first that cited the paper.

Additionally, these high-degree scientists are distributed over the different classes, approximately proportional to the size of the classes, see Figure 6. This shows clear fragmentation of the network, in which each class contains one or more scientists with a very high in-degree who is responsible for carrying Granovetter's hypothesis into 'their' new research area. I will refer to these high in-degree scientists as hub-scientists, or shortly hubs.

Evaluating each class independently, it turns out that the distribution of in-degree within each class roughly follows a power law as well<sup>7</sup>, see Appendix Table 1 for details. Figure 7 graphically shows an example of the power law distribution, of class 7 (the physicists). This power law distribution within the classes suggests that each class can be seen as a community, with his own 'preferred' scientists. The findings of the community structure and the in-degree distribution together, suggest that Granovetter's hypothesis is applied in various scientific communities each with their own narrative or scientific debate. Each community furthermore contains a small number of hub scientists that are largely responsible for carrying Granovetter's hypothesis into their field.

## 4. Evolution of the network in time

This section treats the development of the network. First, the development per community-class is analysed. Thereafter, the dynamics of the overall network is described.

### Development per class

A scientific hypothesis that spread to a large number of research fields can be seen as a scientific innovation, and hence compared to the theory of diffusion of innovations. The main characteristic of diffusion of an innovation are the growth curve and the role of different groups corresponding to this growth. This curve shows slow growth at the start, corresponding to innovators. Thereafter, some early adaptors adapt, some of which have many contacts and/or a high status and can therefore be seen as influencers or opinion leaders. Due to these influencers the innovation spreads to the larger majority. Later on, with most of the population already adopted, the growth decreases.

In this research, the above theory can be tested not only by means of the growth data but also using the in-degree and out-degree that represent how responsible the scientist was for the spread and to what extent the scientist was 'copying' ideas of others. The innovators correspond to scientists who were the first to apply Granovetter's hypothesis to their particular research field, or scientific narrative. They have no or little previous examples, hence their out-degree is low. Thereafter, some prominent scientists pick up the idea and integrate it into their own research. These prominent scientists spread the hypothesis to the majority in their scientific field. Hence their in-degree is high. Scientists that are part of the majority have had many previous examples of the application of Granovetter's hypothesis in their research field. They therefore have a high out-degree. In Section 2 we have already seen that each class contains one or two scientists, hubs, with a very large in-degree. These hubs corresponds to influencers or opinion leaders in the diffusion of innovation terminology.

More formally the hypotheses are as follows. The growth curve in each class will be similar to Roger's logarithmic growth function. The in-degree will show a peak at the point where the hub enters the network. The out-degree is expected to be low in the early years of a cluster, and higher in the later years.

Tracking the number of new nodes per year, exponential growth is found in all classes, see Appendix Table 2. This finding is in line with the growth curve of Roger's innovators-early adaptors-majority hypothesis, though the decreasing growth phase of the late majority is never reached. It is important to note that his exponential growth is visible in classes in different times. This invalidates the

<sup>6</sup> The power law is found using regression on the linear form;  $\log p(x) = \alpha \log x + c$ . I realize that the standard errors are not normally distributed and hence the estimation of fit ( $r^2$ ) does not have much power. For the purposes of this paper, however, this indication of a power law is sufficient and more thorough proof is left for future studies.

<sup>7</sup> Please see footnote 4.

argument that the exponential growth is only due to the increasing number of papers online, hence present in my dataset. In fact, when we consider the growth of papers worldwide according to the Science Citation Index Papers covered, see Appendix Figure 1, we see that the growth of Granovetter's weak-tie-diffusion subfield is quicker than the overall growth of science. This exponential growth pattern per class seems natural for diffusion of a hypothesis that becomes highly popular and therefore further confirms the validity of the partitioning made in the network.

The average out-degree shows an increasing pattern, see Figure 8 and Appendix Figure 2. This means that the first scientists reference few previous scientists that applied Granovetter's hypothesis in their research field. Later scientists use more previous examples. This is in line with my hypothesis and the theory of innovators taking risks and the majority being more like followers.

In most classes the average in-degree shows the following pattern over time, see Appendix Figure 3. The in-degree is relatively low in the first and last years, but experiences one or sometimes multiple peaks in the phase of the early adopters. Splitting these average peaks into all contributing scientists, the amplitude of each peak turns out to be caused by only one or two scientists, namely the hubs of that particular class. See Figure 9 for an example of the evolution of the in-degree.

Putting all these results together, a class is initiated with some innovators that use very few previous examples. The growth of the class is slow, until some hub references to Granovetter's 1973 paper. This creates the first high peak in the in-degree. Next, the growth starts increasing, as does the average out-degree. Some other hubs enter the network and the growth curve steepens further. After the last hub, the in-degree decreases over time. Video 1 shows the growth of class 7 (the physicists) illustrating this process.

#### Dynamics of the overall network

It is hard to specify the evolution of the overall network precisely, but the evolution per class discussed above together with the following findings describe the overall pattern.

The first 17 scientists referencing 'The Strength of Weak Ties' belong to seven distinct classes, as identified in 2012. There are no edges between these first 17 authors, meaning that there are no cross-references between the papers in which they reference Granovetter. After the first 10 years, the network already contains members of eleven, out of the thirteen classes considered, see Figure 10 for details. The presence of members of a class however, does not imply sudden growth of that class in the nearby future. As described earlier, almost all classes start with innovators and the growth of the class is exponential which means that the first growth is slow.

The first classes that grow considerably are the social classes, red (1) /bright green (16) /purple (9) in Figure 2. Thereafter, economics & business classes, blue (5) /light blue (2) /bright orange (11) /dull orange (15) /dark green (8) in Figure 2, start growing as well. Roughly in the 1990s computer scientists of class 4 and 13, yellow in Figure 2, start referencing to Granovetter. Around 2000 'The Strength of Weak Ties' is picked up on by the physicists, class 7 - black in Figure 2.

In sum, the interesting result is that the first scientists who use Granovetter's hypothesis, both in the overall network and for each class individually, are generally not connected to each other. The hubs that enter the classes somewhat later than these innovators, are the ones that initiate cluster-forming; 'the majority' references these hubs and thereby creates a cluster around the hubs.

The findings of the developments within each class and the development of the network overall confirm the previous finding that Granovetter's hypothesis is applied in various scientific communities each with their own narrative and hubs. In addition, the hubs can be seen as 'opinion leaders' that enable the hypothesis to spread to the majority. The hubs are crucial for the clustering of the network, which is shown to be formed around these hubs.

## 5. Close-up: Hubs

The previous analysis reveals importance of the hubs. In fact, it seems as if each class can be represented by their hubs. In the analysis of the community structure, four classes were found that I broadly described as social science classes. Although these classes contain researchers of approximately the same research areas, my hypothesis was that each class fits Granovetter's hypothesis into a different narrative, a different debate. To see whether this is the case and simultaneously investigate if classes can indeed be represented by their hubs, I will describe the differences between the classes that I broadly described as social science classes based on the most-cited papers of the hubs. Thereafter I test some simple hypothesis to gain insight into the question: What makes a scientist a hub?

#### Social Science Classes

I identified four classes that I broadly categorized as social science classes because these classes all contain significantly more social scientists and less business & economists than the overall network. These classes are class 1, red in Figure 2, class 16, bright green in Figure 2, class 9, purple in Figure 2, and class 12, yellow-green in Figure 2. From a narrative perspective, I expect that these classes use Granovetter's hypothesis for different debates each with their own narrative. In the following I will show how these narratives can broadly be determined. The method I used was as follows; for each of these classes I read the articles that were most often referenced in this network of the two largest hubs. This thus concerns the articles that were referenced by new authors that used Granovetter's hypothesis for the first time, hence not the articles that are most often referenced in general. Next, I gathered the titles of all the papers and books as a response to these most-referenced articles. Again these titles concern papers or books by authors that use Granovetter's hypothesis for the first time.

*Class 1.* The hub-scientists of class 1 that I evaluated are Peter van Marsden, currently Dean of Social Science at Harvard University, and Nan Lin, currently Professor of Sociology at Duke University. Marsden's most referenced articles in this network are 'Measuring Tie Strength' (1984), 'Core Discussion Networks of Americans' (1987) and 'Network Data and Measurement' (1990). The most referenced papers of Nan Lin are 'Social Resources and Strength of Ties: Structural Factors in Occupational-Status Attainment' (1981), and 'Access to Occupations through Social Ties' (1986).

Based on the content of these articles and by counting the occurrence of particular words in the titles of responses to these articles, I distinguished the following themes around these hubs: 'the job-market', 'health & well-being', 'social support/ social resources', and 'social capital', see Appendix Table 4. In these titles, the term social capital is mostly used in conjunction with words from one of the other themes.

*Class 9.* The hub-scientist of class 9 is Barry Wellman, currently Professor of Sociology at the University of Toronto. His most cited articles in this network are 'Different Strokes from Different Folks- Community Ties and Social Support' (1990) and 'Community Question- Intimate Networks of Yorkers' (1979).

The unifying theme in Wellman's articles and the papers and books that came as a response is 'community', with the word 'community' present in the title of in 21% of all responses. Otherwise, the subjects covered by the response papers are not easily classifiable. This observation corresponds to the location of scientists of class 9, in Figure 2, which do not form a dense cluster but are rather spread out.

*Class 12.* In class 12, the two hubs-scientist are Michael Woolcock, currently both Lead Social Development Specialist in the World Bank's Development Research Group and Lecturer at the Harvard's University's Kennedy School of Government and Deepa Narayan, currently an independent international poverty and

development advisor with experience at the UN and the WorldBank. Woolcock's most cited article is *'Social Capital and Economic Development: Toward a theoretical Synthesis and Policy Framework'* (1998). In 1990, Woolcock and Deepa Narayan published a paper together, called *'Social Capital: Implications for Development Theory, Research and Policy'*, that is the second most referenced paper of class 12.

As already apparent from the paper-titles and positions of Woolcock and Narayan, their main focus is social-economic development and policy, in which they claim social capital plays a key role (Woolcock & Narayan, 2000). The term 'social capital' is used in 26% of titles of responses to the articles of Woolcock and Narayan. In addition, almost all titles refer to a specific region or country, which often correspond to a rural area.

**Class 16.** Of class 16 I evaluated two authors; Ronald Steward Burt, currently Professor of Sociology and Strategy at University of Chicago Booth School of Business, and Scott Archer Boorman, currently Professor of Sociology at Yale. The most referenced paper of Burt is *'Network Items and the General Social Survey'*. Boorman's most referenced paper is *'Social-Structure from Multiple Networks .1. Blockmodels of Roles and Positions'*.

The commonality between the articles of Burt and Boorman and the response-papers and books is the theoretic-analytical approach to social networks. Class 16's interest is the structure and dynamics of networks, where titles frequently contain terms such as 'closure', 'structure', 'analysis' 'dynamics' and 'bridge/bridging', see Appendix Table 5. Figure 2 also shows that class 16 is closest connected to class 7 (the physicists) who share the same analytical interest. Taking a second look at the distribution of research areas in class 16, it is found that this class indeed contains relatively many scientists of the areas mathematics and mathematical methods in social science.

In sum, all these classes seem to speak of a different scientific debate. From their research areas, as identified by Web of Science, the classes seem similar. Network analysis of the research of this paper, however, showed that these classes can be divided into four groups and this short investigation of the hubs confirms their separateness by the scientific topics scientists of these classes write about. Hence, with the small research above on the most frequently referenced papers of hubs, it is possible to broadly determine the more specific themes in which scientists use Granovetter's hypothesis. This is not to say that the field characterisation I made is hundred per cent accurate, since the research was limited, but together with the scientists' research areas, location and connection between the classes the characterization seems valid for now.

#### **What makes a scientist a hub?**

Why are some scientists so often referenced in this network, serving as a representative in their class for Granovetter's hypothesis? We have already seen that being among the first scientists of a class to reference Granovetter is not the answer. The hubs rather enter slightly later, after the first 'unsuccessful' innovators. I formulate four hypotheses that can help to answer this question and make an effort to test two of them.

First, a simple hypothesis is that the hub-scientists are explorative; they did not copy Granovetter's idea from many others, but they found out about the hypothesis themselves. Second, an alternative simple hypothesis is that hubs are scientists that form a bridge between to communities of the underlying network. Third, the hubs could be well-known scientists in their field, even before the use of Granovetter's hypothesis. Finally, it could be that the hub-scientists were successful in fitting Granovetter's hypothesis into the narrative of their subfield; although some innovative scientists of their class referenced Granovetter earlier, their interpretation and use of his hypothesis was perhaps not fruitful.

The relation between out-degree and in-degree does not show a straightforward pattern: there are many explorative authors with a low out-degree, but only few of them became hubs. Some

hubs, on the other hand, have a relatively high out-degree. These two facts both contradict the first hypothesis.

The second hypothesis, that hubs are often bridges between communities, is not confirmed either by the data. By identifying the classes of the scientists that the hubs referenced we find some hubs that mainly reference to scientists of their own class. Moreover, there are innovators that reference only to scientist from other classes. Therefore, being a bridge can never be a sufficient condition for being a hub.

The third hypothesis, that hubs are prominent scientists in their subfield, is not something that can be concluded from this data, since the ties and node size in this network are only based on references around Granovetter's hypothesis. Fortunately, this hypothesis is possible to test using data on citations of scientists in general and I plan to look further into this in future research.

To determine whether the hub scientists were better than the innovators in fitting Granovetter's hypothesis into the narrative of their subfield a more thorough qualitative analysis is needed. This was unfortunately beyond the scope of this research.

## **6. Conclusion**

Spread of science is a requirement for scientific progress, but how does a scientific theory spread? This research tracked the spread of one particular scientific hypothesis as a case study for the diffusion of science. The data suggest that the following three issues determine the diffusion pattern of scientific hypotheses.

First, the diffusion of a new scientific hypothesis is a process that involves conceptualisation and interpretation. The way in which the hypothesis is interpreted depends on the key narrative of a given scientific community. As a result, the scientists who use the hypothesis under consideration are grouped into communities representing different debates or key-narratives. The number of references within these communities is relatively large, and there are relatively few cross-references between the communities.

Second, the diffusion into a new scientific field is similar to the diffusion of an innovation. There are innovators that apply the scientific hypothesis first. After a period of slow growth, some early adopters use the hypothesis in their research and they quickly spread the hypothesis to the majority. This pattern is confirmed by the exponential growth within the communities. Moreover, the data of this case study show that the first scientists applying the hypothesis in a new scientific field, use little previous examples. They reference few other scientists who have already used the hypothesis. Scientists that start applying the hypothesis later, on the other hand, use many previous examples. This increase in the number of previous examples further characterises early scientists as innovators and later scientists as followers.

Third, due to the reference behaviour of scientists, communities are formed around the hubs of each particular research field. As a result, the different debates to which the hypothesis has spread or led can be represented by the hubs of that field.

In particular, this research determined the general narrative of four classes that, based on their research areas, seemed to be operating in the same scientific field. By focussing on the most referenced articles of the hubs and the titles of the responses to these articles, the key topics of research were quickly determined for each class, distinguishing them from one another. This analysis technique could be therefore also be useful for the writing of literature reviews or in determining the impact of a scientific paper.

Finding these three mechanisms raises interesting questions for diffusion and scientific progress. Firstly, one could ask whether these mechanisms foster or hinder diffusion. In case diffusion is hindered, opportunities for improving scientific spread arise simultaneously. In case diffusion is stimulated, this knowledge can be used for diffusion in a different but similar context. For example, an interesting analogy can be drawn between the scientific community and society. As this research confirms, the scientific

community consists of multiple communities each with their own topics of debate and corresponding key-narrative. Society analogously consist of different groups of people, each holding particular conversations andhaving their own world view. The way an idea or belief, such as a political belief, spreads within society can hence be compared to the diffusion of a scientific theory. Insights into the processes behind the diffusion of science can therefore be valuable for politicians and society as a whole.

## References

- Barabási, A. L., Albert, R., & Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, 281(1), 69-77.
- Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science*, 15, 215-227.
- Bazerman, C. (1990). Discourse analysis and social construction. *Annual Review of Applied Linguistics*, 11, 77-83.
- Bird, A. (2007). What is scientific progress?. *Noûs*, 41(1), 64-89.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
- Burt, R. S. (1984). Network items and the general social survey. *Social networks*, 6(4), 293-339.
- Burt, R. S. (2000). The network structure of social capital. *Research in organizational behavior*, 22, 345-423.
- Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996), 1194-1197.
- Centola, D., Eguíluz, V. M., & Macy, M. W. (2007). Cascade dynamics of complex propagation. *Physica A: Statistical Mechanics and its Applications*, 374(1), 449-456.
- Chow, G. C. (1967). Technological change and demand for consumers. *American Economic Review*, 57, 1117-1130.
- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4), 661-703.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology*, S95-S120.
- Dezsó, Z., & Barabási, A. L. (2002). Halting viruses in scale-free networks. *Physical Review E*, 65(5), 055103.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB Journal*, 22(2), 338-342.
- Floyd, A. (1962). Trend forecasting: A methodology for figure of merit. In J. Bright (Ed.), *Technological forecasting for industry and government* (pp. 95 - 105). New Jersey: Prentice Hall.
- Fourt, L. A., & Woodlock, J. W. (1960, October). Early prediction of early success of new grocery products. *Journal of Marketing*, 25, 31-38.
- Goldenberg, J., Han, S., Lehmann, D. R., & Hong, J. W. (2009). The role of hubs in the adoption process. *Journal of Marketing*, 73(2), 1-13.
- Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 1360-1380.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O., & Peacock, R. (2005). Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Social science & medicine*, 61(2), 417-430.
- Kajikawa, Y., Ohno, J., Takeda, Y., Matsushima, K., & Komiyama, H. (2007). Creating an academic landscape of sustainability science: an analysis of the citation network. *Sustainability Science*, 2(2), 221-231.
- Karsai M., Kivela M, Kaski K, Barabási A.L, et al. (2011), Small but slow world: How Network topology and burstiness slow down spreading, *American Physical Society, physical review E* 83, pp.360-364
- Lin, N. (1999). Social networks and status attainment. *Annual review of sociology*, 467-487.
- Mansfield, E. (1961). Technical change and the rate of imitation. *Econometrica*, 29, 741-766.
- Marsden, P. V. (1987). Core discussion networks of Americans. *American sociological review*, 122-131.
- Marsden, P. V., & Campbell, K. E. (1984). Measuring tie strength. *Social forces*, 63(2), 482-501
- Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation—A 25-year review. *International Journal of Forecasting*, 22(3), 519-545.
- Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of management review*, 23(2), 242-266.
- Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5200-5205.
- Phelps, Corey, Ralph Heidl, and AnuWadhwa. "Knowledge, networks, and knowledge networks a review and research agenda." *Journal of Management* 38.4 (2012): 1115-1166.
- Price, D. D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5), 292-306.
- Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative science quarterly*, 48(2), 240-267.
- Rogers, E. M. (1962). *Diffusion of innovations*. New York: The Free Press.
- Tarrow, S., & Tollefson. (1994). *Power in movement: Social movements, collective action and politics* (pp. 3-4). Cambridge: Cambridge University Press.
- Tornberg P. & Tornberg A., (2014), *Modeling Free Social Spaces – Diffusion of (Complex) Contention*, not yet published
- Valente, T. W., & Rogers, E. M. (1995). The origins and development of the diffusion of innovations paradigm as an

example of scientific growth. *Science communication*, 16(3), 242-273.

Wagner, C. S., & Leydesdorff, L. (2005). Mapping the network of global science: comparing international co-authorships from 1990 to 2000. *International Journal of Technology and Globalisation*, 1(2), 185-208.

Wellman, B. (1979). The community question: The intimate networks of East Yorkers. *American journal of Sociology*, 1201-1231.

Wellman, B., & Wortley, S. (1990). Different strokes from different folks: Community ties and social support. *American journal of Sociology*, 558-588.

White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American journal of sociology*, 730-780.

Zaheer, A., & McEvily, B. (1999). Bridging ties: A source of firm heterogeneity in competitive capabilities. *Strategic management journal*, 20(12), 113



## Figures

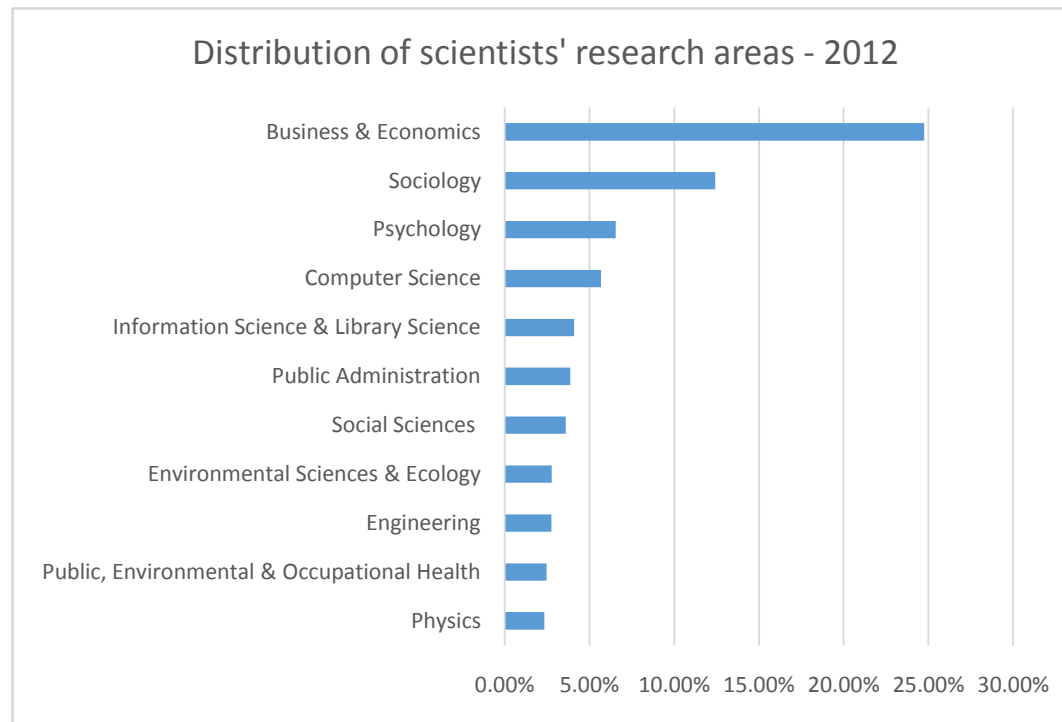
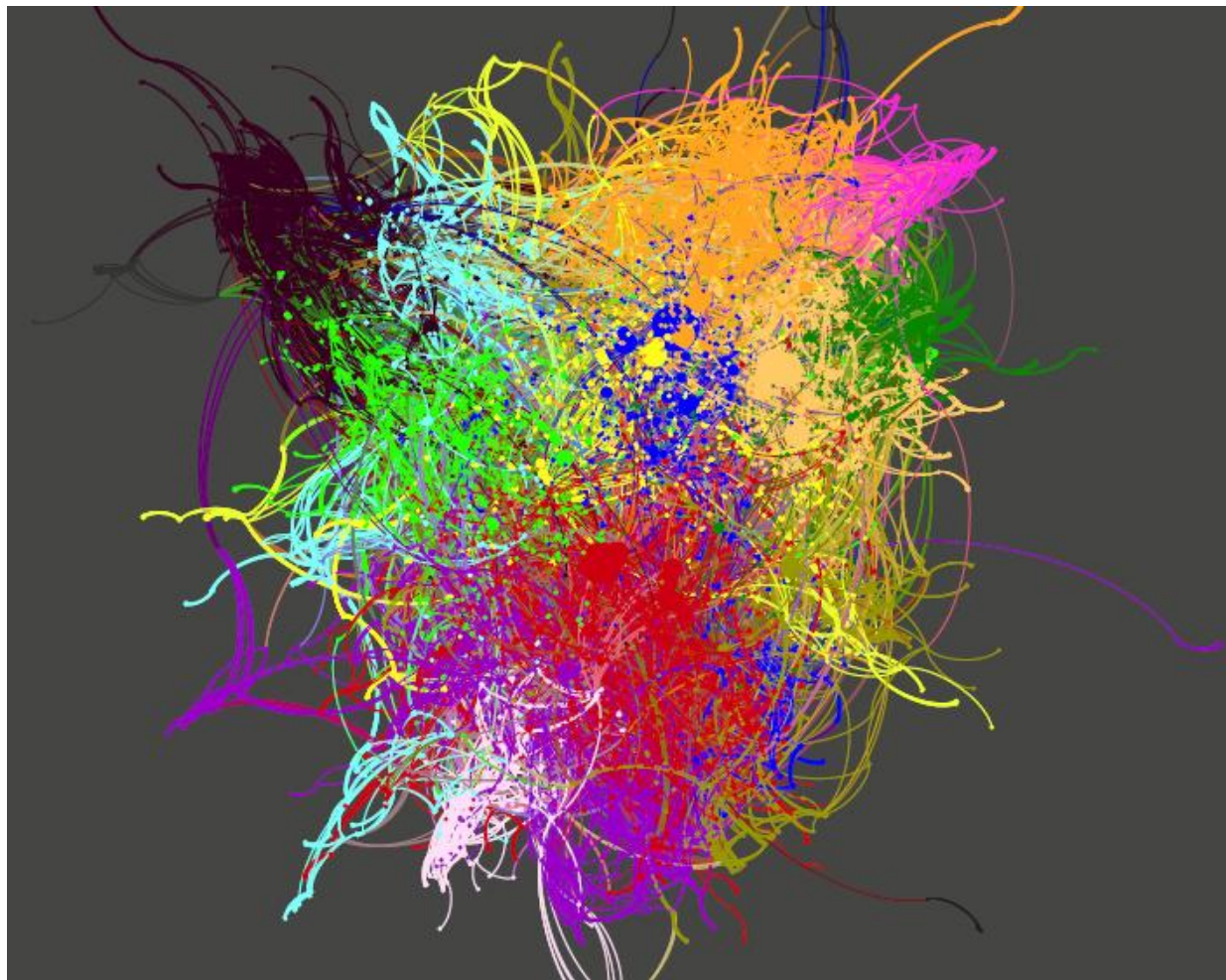


Figure 1: This figure shows the distribution of scientists' research areas in the network in 2012. Only the largest 11 research areas are shown.



1	Social scientist core	(18.58%)
11	Business & Economics Core	(12.75%)
16	Social scientists group 2	(8.91%)
5	Business & Economics group 2	(8.7%)
9	Social scientists group 3	(8.55%)
15	Business & Economics group 3	(8.28%)
7	Physicists, Science & Technology	(7.51%)
2	Business & Economics group 4	(5.69%)
4	Computer Scientists	(5.35%)
8	Business & Economics group 5	(4.99%)
12	Social scientist group 4	(4.96%)
10	Government, Law & communication scientists	(2.08%)
13	Computer Scientists	(1.6%)
0	Geographers	(1.33%)

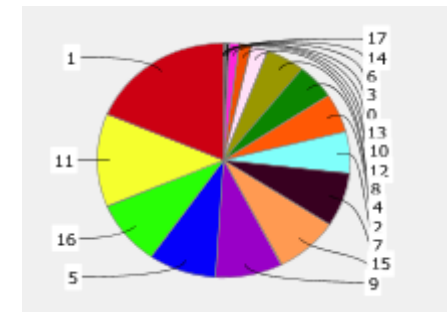
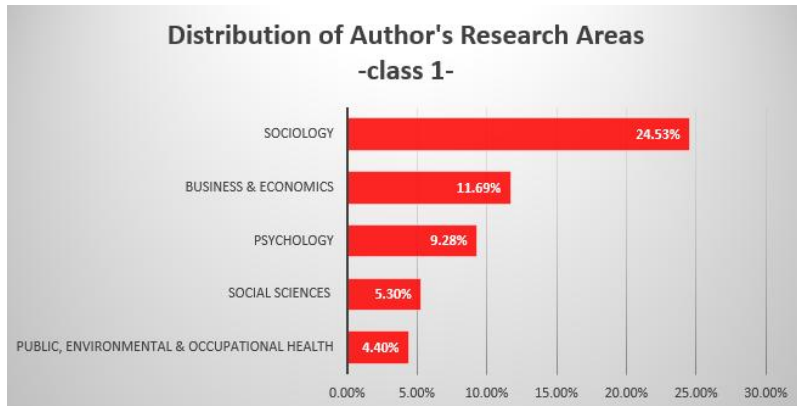


Figure 2: This network shows all 6296 scientists that all play some role or another in the diffusion process as nodes. The colours correspond to the different clusters found in the network. The legend only shows the colours of the 13 largest clusters in the network.

a)



b)

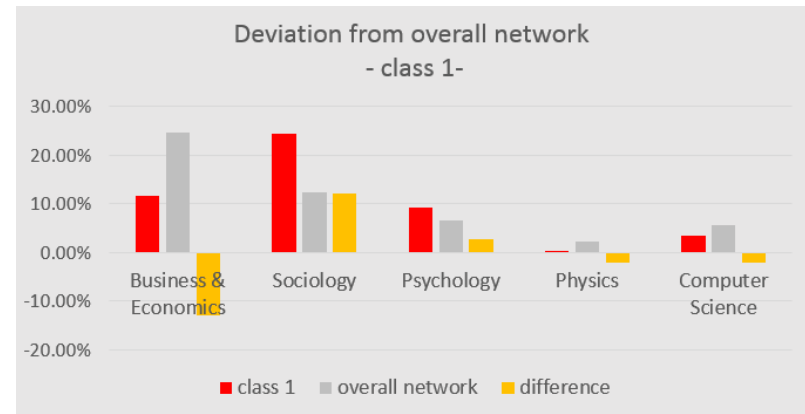
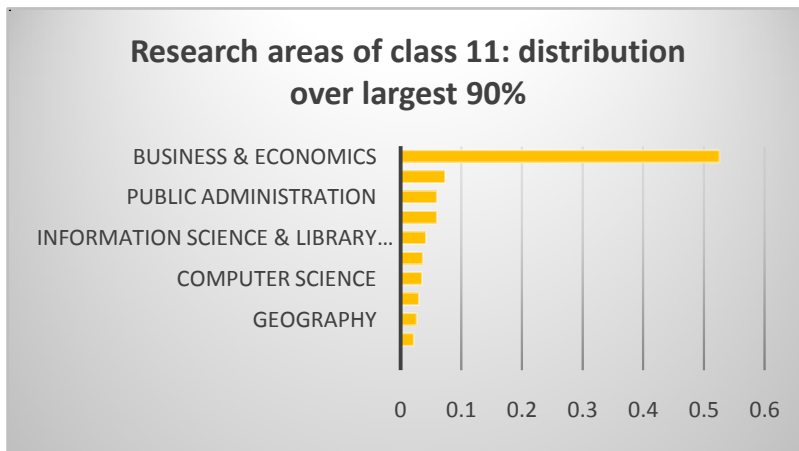


Figure 3:

a) This figure shows the distribution of scientists of class 1 over the five largest research areas. b) This figure compares the distribution of class 1 with (red) with the overall network (grey). Only the research areas with the largest deviations (yellow) are shown.

a)



b)

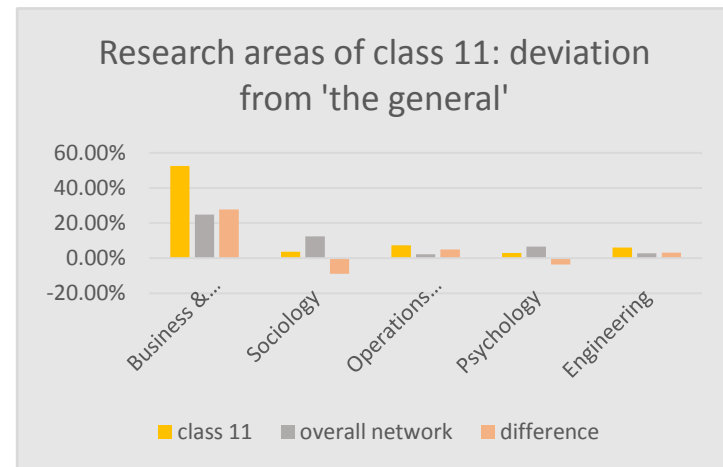


Figure 4: a) This figure shows the distribution of scientists of class 11 over the five largest research areas. B) This figure compares the distribution of class 11 with (yellow) with the overall network (grey). Only the research areas with the largest deviations (pink) are shown.

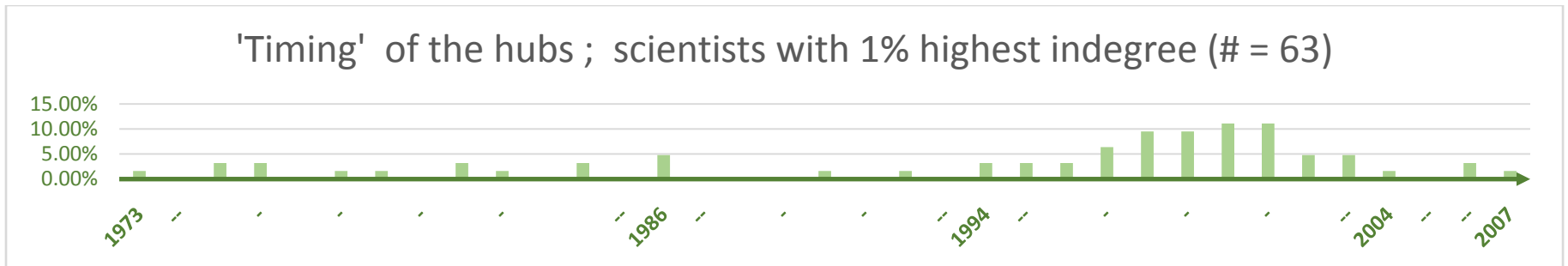


Figure 5: This timeline shows which percentage of the highly referenced scientists, 'hubs', referenced Granovetter in each year. Most important from this timeline is that the hubs did not all enter in the first years after Granovetter's publication, but are spread over the period 1973-2007.

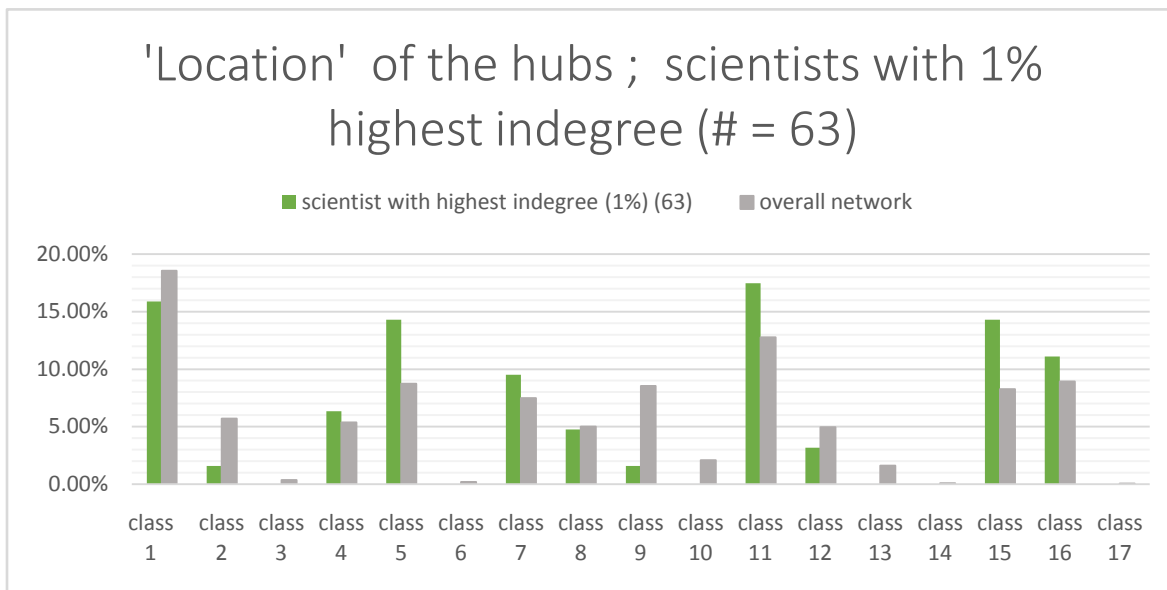


Figure 6: This figure shows which percentage of the highly reference scientists, 'hubs', are present in each class. The figure demonstrates that the hubs are distributed over the different classes, approximately proportional to the size of the classes.

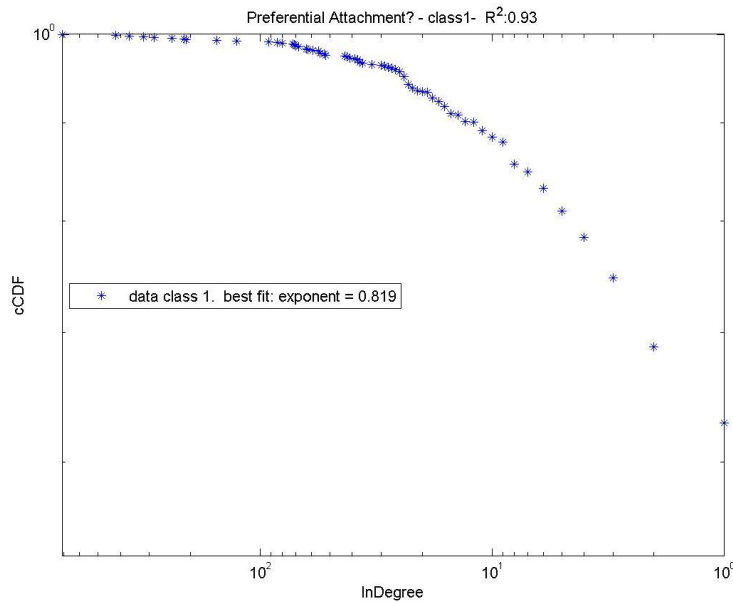
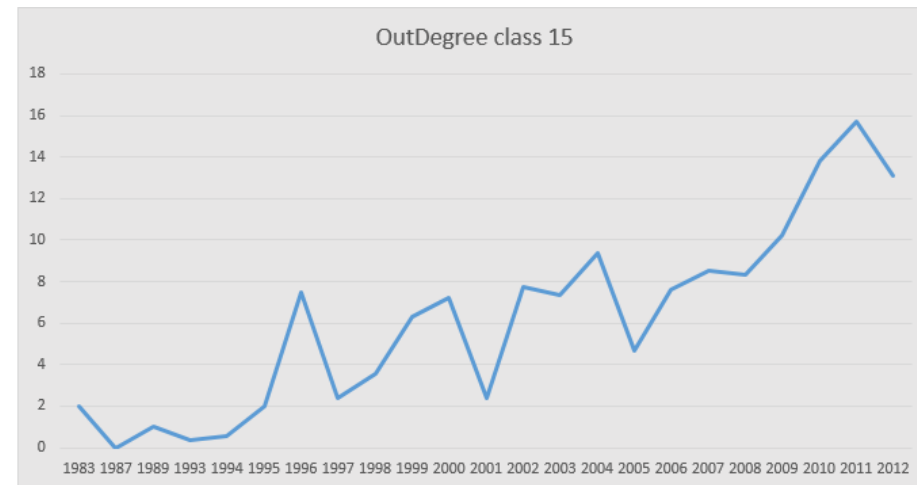


Figure 7 : This figure shows the cumulative distribution for the in-degree of class 1. This plot demonstrates that the in-degree seems to follow a power law with exponent approximately equal to 0.8. This kind of uneven distribution is a sign of preferential attachment, otherwise known as cumulative advantage, which refers to the fact that popular scientists are more likely to be referenced again creating an uneven distribution w the majority is only rarely referenced and a small amount of scientists is referenced disproportionately often.

Figure 8: This figure shows the average out-degree over time of class 15. Class 15 is a community with mainly economics & business scientists. The out-degree is clearly rising. This demonstrates that the first scientists of class 15 did not use many previous examples for the use of Granovetter's hypothesis in their research field. This characterizes them as followers. Later scientists use increasingly more examples, characterizing them more as followers. A similar increasing pattern of the out-degree is found in the other classes.



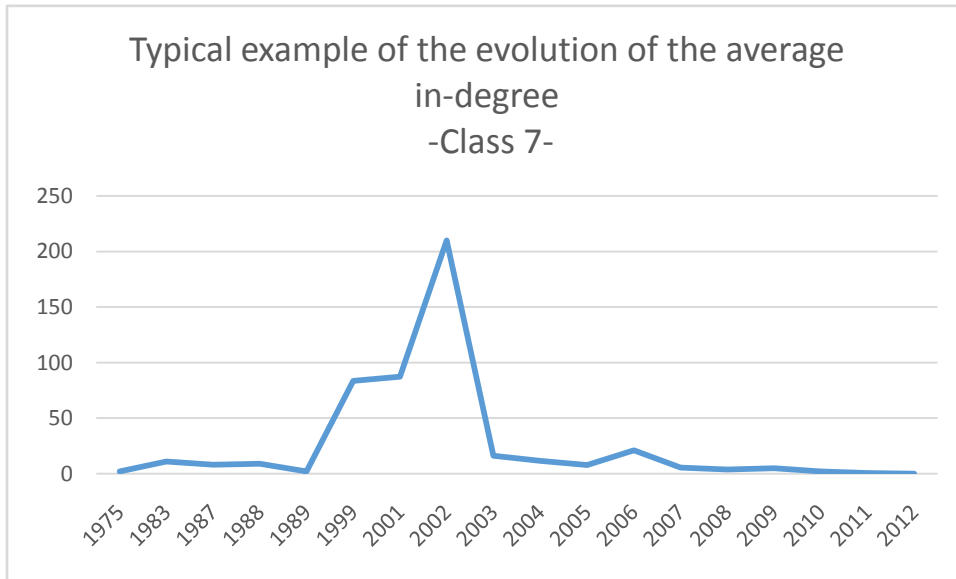
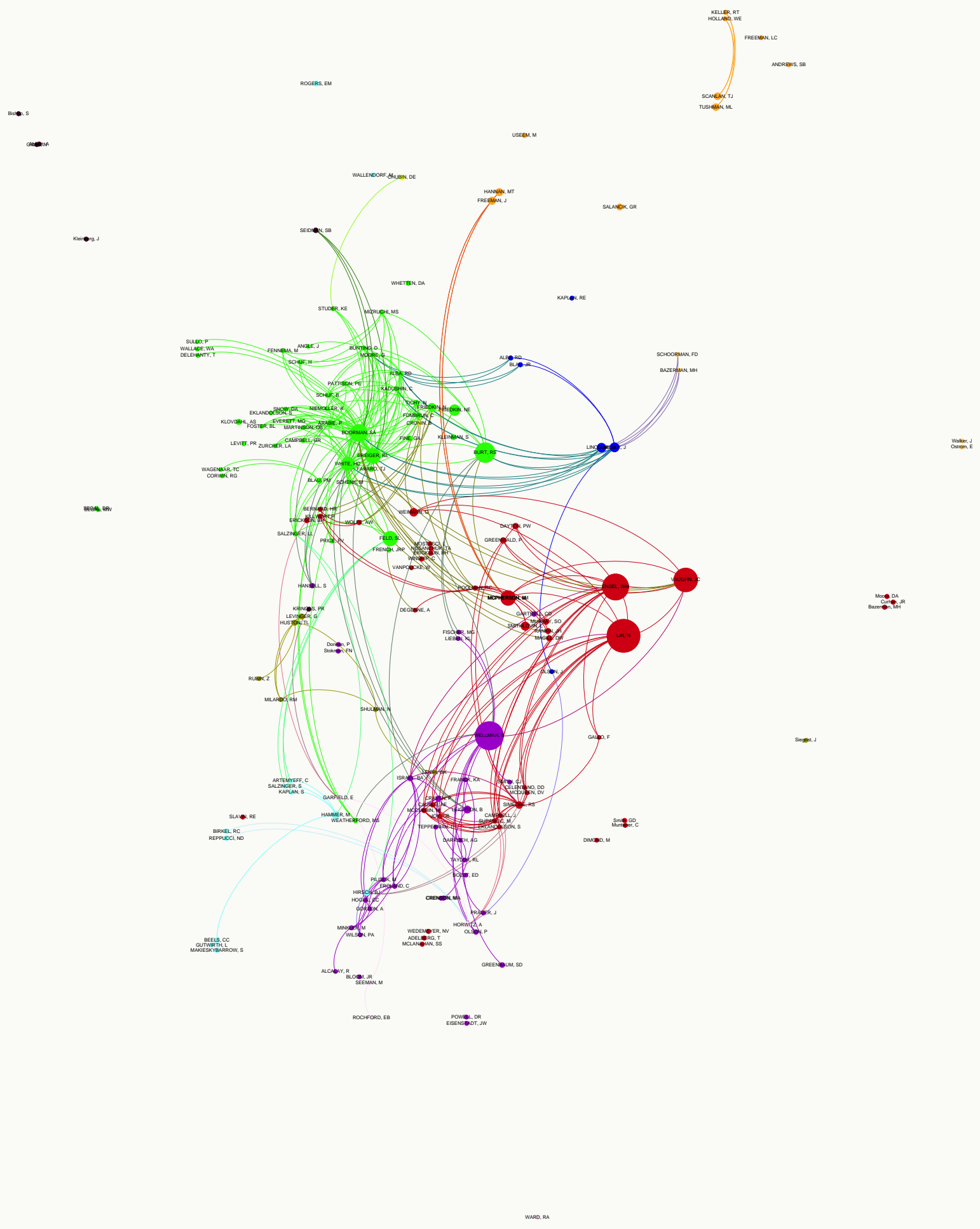


Figure 10: This figure shows average in-degree of class 7 over time. The pattern of a low average in-degree in the first and last years with some sudden high spikes in between is found in other classes as well. The amplitude of the peaks is caused by the hub-scientists who are publishing their highly referenced papers in those years.

Figure 10 (See following page): This network shows all the nodes that scientists that have referenced to Granovetter's 1973 paper before 1984. This network shows that scientists of already 11 out of the 13 clusters have referenced Granovetter. This illustrates that Granovetter's hypothesis did not first spread within one cluster and thereafter diffused to another clusters. Rather, it shows that there were innovators of multiple scientific fields that used Granovetter's hypothesis in their research. The network further shows that the social science clusters, red, bright-green, purple and yellow-green, were the first to grow considerably.





## Appendix

Power law distribution within classes														
class	0	1	2	4	5	7	8	9	10	11	12	13	15	16
exponent	0.29	0.82	0.85	0.49	0.61	0.61	0.58	0.79	0.59	0.74	0.78	0.61	0.52	0.76
r <sup>2</sup>	0.44	0.93	0.68	0.65	0.69	0.68	0.78	0.67	0.28	0.76	0.71	0.25	0.49	0.94

Table 1. This table shows the results from fitting the in-degree distribution of each class to a power law. The first figure gives the exponent and the second figure gives the R<sup>2</sup> value which indicates the fit. R<sup>2</sup> lies between 0 and 1, where 1 is a perfect fit. This is by no means a 'proof' that the data fits a power law. It is rather an indication that the data follows a power law. In class 0, class 10 and class 13 the fit is poor. This could be due to the size of these classes; these classes contain less than 2 per cent of all nodes, whereas all other classes under consideration contain more than 5%.

Exponential Growth within classes														
class	0	1	2	4	5	7	8	9	10	11	12	13	15	16
exponent: 10 <sup>slope</sup>	1.14	1.11	1.08	1.16	1.14	1.16	1.17	1.09	1.05	1.15	1.11	1.07	1.18	1.07
R <sup>2</sup>	0.39	0.91	0.75	0.78	0.92	0.93	0.84	0.82	0.51	0.95	0.86	0.54	0.88	0.76

Table 2. This table shows the results from fitting the growth in each class to exponential growth. The first figure gives the exponent and the second figure gives the R<sup>2</sup> value which indicates the fit. R<sup>2</sup> lies between 0 and 1, where 1 is a perfect fit. In the classes 0, 10 and 13 the R<sup>2</sup> is low, but this could be explained by their size; these classes contain less than 2 per cent of all nodes, whereas all other classes under consideration contain more than 5%.

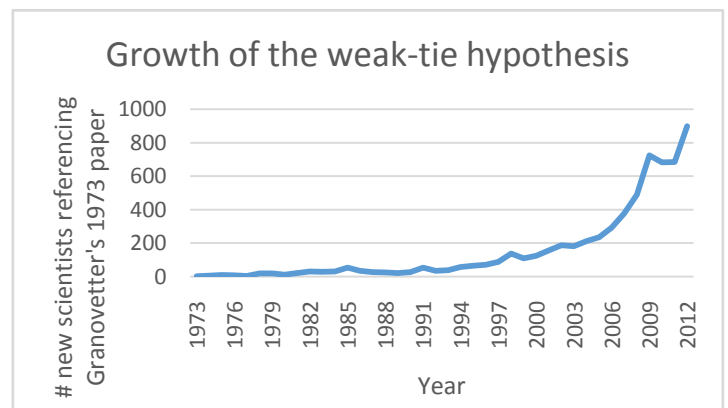
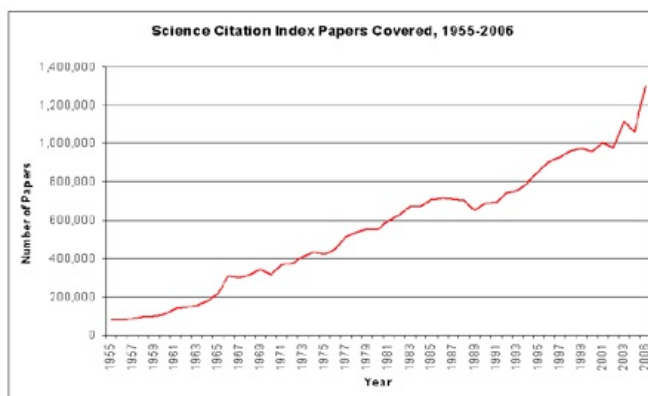


Figure 1. The Figure on the right shows for each year the amount of scientists that start referencing to 'The Strength of Weak Ties'. Each year hence excludes scientists that have already referenced the paper earlier. The Figure shows exponential growth. The Figure on the left shows the evolution of the total number of published papers. Source: Barabási, A.L, Barzel B., Martino M, 'Network Science', February 2012, Available at <http://www.barabasilab.com/>; [barabasilab.neu.edu/courses/.../Class7.../07\\_CLASS\\_2012\\_BAmodel.ppt](http://barabasilab.neu.edu/courses/.../Class7.../07_CLASS_2012_BAmodel.ppt)

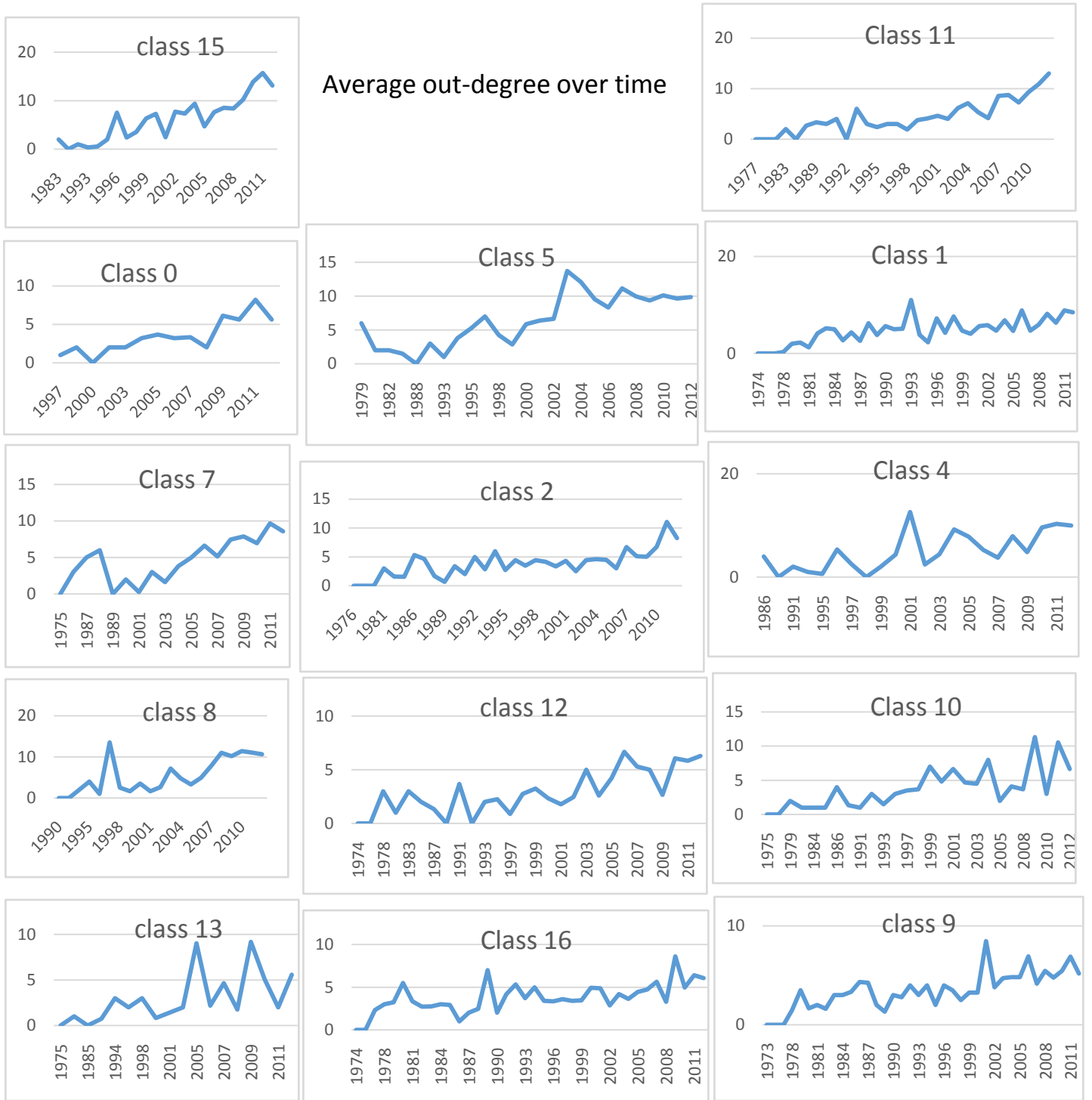


Figure 2. This Figure shows for each class the average out-degree over time. The general conclusion from these graphs is that the out-degree increases over time.

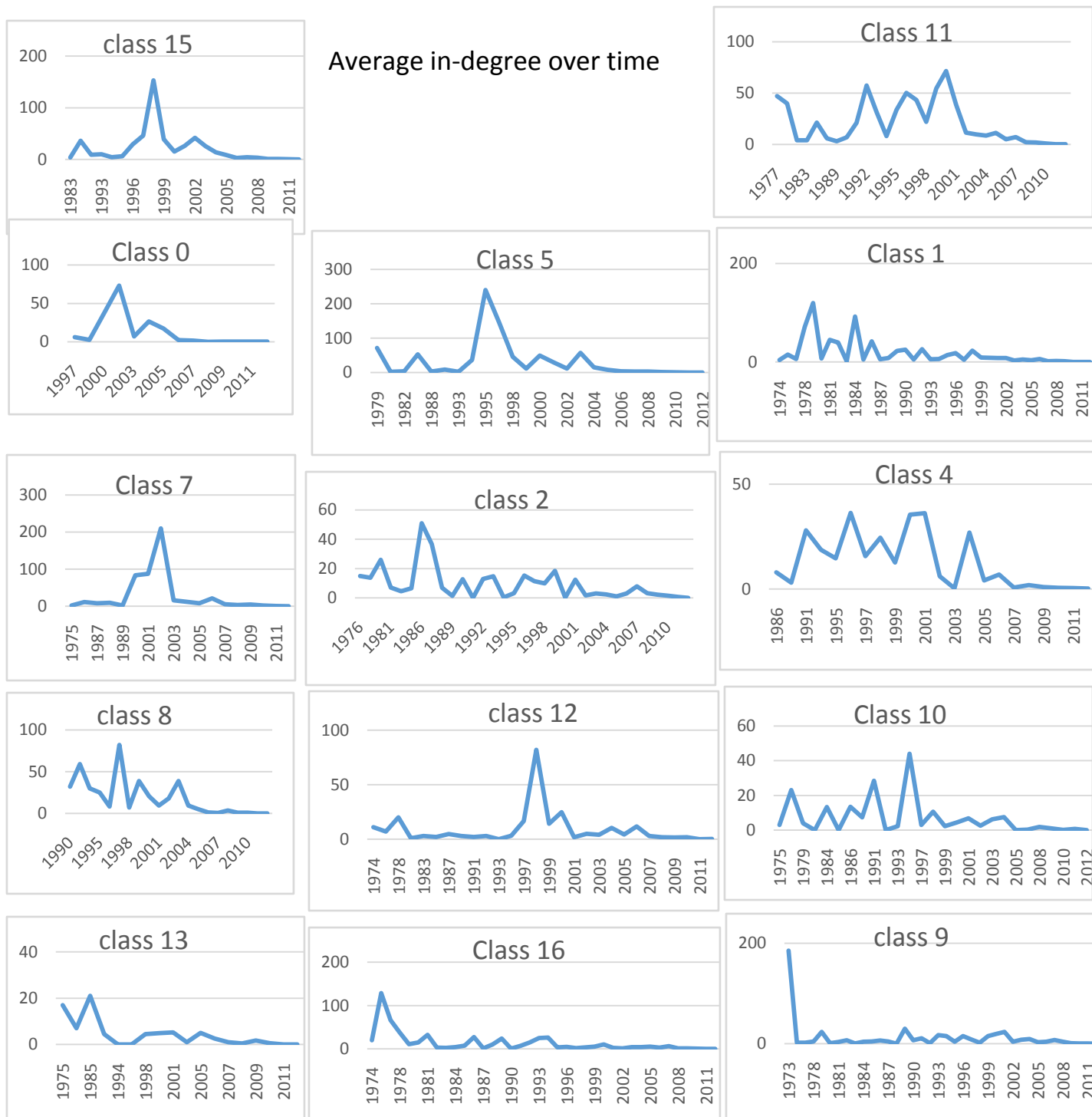


Figure 3. This Figure shows for each class the average in-degree over time. In most classes, the in-degree is low in the first and last years, and experiences one or several peaks in the middle years.

theme	count	% of total
jobs	<b>499</b>	<b>23.31</b>
labour	111	5.18
employment	42	1.96
job	155	7.24
career	41	1.91
entrepreneur	48	2.24
manager	23	1.07
occupations	43	2.01
college	12	0.56
school	24	1.12
social resources -		
social support	<b>156</b>	<b>7.29</b>
social resources	78	3.64
social support	78	3.64
inequalities	<b>382</b>	<b>17.84</b>
equality	35	1.63
status	106	4.95
gender	68	3.18
ethnic	42	1.96
race	52	2.43
urban	54	2.52
rural	25	1.17
health	<b>167</b>	<b>7.80</b>
mobility	41	1.91
health	67	3.13
well being	14	0.65
depression	11	0.51
stress	34	1.59
social capital	<b>345</b>	<b>16.11</b>

Table 4. This table shows the occurrence of words in the titles of articles and books that referenced to the most cited articles by Peter van Marsden [*'Measuring Tie Strength'* (1984), *'Core Discussion Networks of Americans'* (1987) and *'Network Data and Measurement'* (1990)] and Nan Lin [*'Social Resources and Strength of Ties: Structural Factors in Occupational-Status Attainment'* (1981), and *'Access to Occupations through Social Ties'* (1986)].

theme	count	% of total
Network Analysis	<b>128</b>	<b>43.84%</b>
bridge/bridging	10	3.42%
structure	35	11.99%
analysis	30	10.27%
dynamics	12	4.11%
holes	10	3.42%
closure	13	4.45%
data	18	6.16%

Table 5. This table shows the occurrence of words in the titles of articles and books that referenced to the most cited articles by Ronald Steward Burt [ *'Network Items and the General Social Survey'*] and Scott Archer Boorman [ *'Social-Structure from Multiple Networks .1. Blockmodels of Roles and Positions'*].