

**A Few Special Cases:  
Scientific Creativity and Network  
Dynamics in the Field of Rare Diseases**

*M. Laura Frigotto, Massimo Riccaboni*





**DISA**

**Dipartimento di Informatica  
e Studi Aziendali**

**2011/3**

**A Few Special Cases:  
Scientific Creativity and Network  
Dynamics in the Field of Rare Diseases**

*M. Laura Frigotto, Massimo Riccaboni*



**UNIVERSITÀ DEGLI STUDI  
DI TRENTO**

**DISA WORKING PAPER**

## **DISA Working Papers**

The series of DISA Working Papers is published by the Department of Computer and Management Sciences (Dipartimento di Informatica e Studi Aziendali DISA) of the University of Trento, Italy.

### **Editor**

Ricardo Alberto MARQUES PEREIRA    ricalb.marper@unitn.it

### **Managing editor**

Roberto GABRIELE                            roberto.gabriele@unitn.it

### **Associate editors**

Michele ANDREAUS	michele.andreaus@unitn.it	Financial and management accounting
Flavio BAZZANA	flavio.bazzana@unitn.it	Finance
Pier Franco CAMUSSONE	pierfranco.camussone@unitn.it	Management information systems
Luigi COLAZZO	luigi.colazzo@unitn.it	Computer Science
Michele FEDRIZZI	michele.fedrizzi@unitn.it	Mathematics
Andrea FRANCESCONI	andrea.francesconi@unitn.it	Public Management
Loris GAIO	loris.gαιο@unitn.it	Business Economics
Umberto MARTINI	umberto.martini@unitn.it	Tourism management and marketing
Pier Luigi NOVI INVERARDI	pierluigi.noviinverardi@unitn.it	Statistics
Marco ZAMARIAN	marco.zamarian@unitn.it	Organization theory

### **Technical officer**

Mauro MION                                    mauro.mion@unitn.it

### **Guidelines for authors**

Papers may be written in English or Italian but authors should provide title, abstract, and keywords in both languages. Manuscripts should be submitted (in pdf format) by the corresponding author to the appropriate Associate Editor, who will ask a member of DISA for a short written review within two weeks. The revised version of the manuscript, together with the author's response to the reviewer, should again be sent to the Associate Editor for his consideration. Finally the Associate Editor sends all the material (original and final version, review and response, plus his own recommendation) to the Editor, who authorizes the publication and assigns it a serial number.

*The Managing Editor and the Technical Officer ensure that all published papers are uploaded in the international RepEc public-action database. On the other hand, it is up to the corresponding author to make direct contact with the Departmental Secretary regarding the offprint order and the research fund which it should refer to.*

### **Ricardo Alberto MARQUES PEREIRA**

*Dipartimento di Informatica e Studi Aziendali*

*Università degli Studi di Trento*

Via Inama 5, TN 38122 Trento ITALIA

Tel +39-0461-282147 Fax +39-0461-282124

E-mail: ricalb.marper@unitn.it

**A Few Special Cases:  
Scientific Creativity and Network Dynamics in the Field of Rare Diseases**

M. Laura Frigotto

Massimo Riccaboni<sup>1</sup>

Department of Management and Computer Science  
University of Trento  
Via Inama, 5 – 38100 Trento

**Abstract**

We develop a model of scientific creativity and test it in the field of rare diseases. Our model is based on the results of an in-depth case study of the Rett syndrome. Archival analysis, bibliometric techniques and expert surveys are combined with network analysis to identify the most creative scientists. First, alternative measures of generative and combinatorial creativity are compared. Then, we generalize our results and present a stochastic model of socio-semantic network evolution. The model predictions are tested with multiple networks of rare disease specialties. We find that new scientific collaborations among experts in a field enhance combinatorial creativity. Instead, high entry rates of novices are negatively related to generative creativity. By extending the set of useful concepts, creative scientists gain in centrality. At the same time, by increasing their centrality in the scientific community, scientists can replicate and generalize their results, thus contributing to a scientific paradigm.

**Keywords:** Knowledge creation, Scientific Networks, Team Formation, Rare Diseases, Network Dynamics, Biomedical research.

**MSC Codes:** 05C82, 05C90, 05C65, 62P25, 91D30, 91B82

**JEL Codes:** C63, L14, L26, L65, O31, O33

---

<sup>1</sup> Corresponding author: Massimo Riccaboni (email: [massimo.riccaboni@unitn.it](mailto:massimo.riccaboni@unitn.it)). The authors wish to thank Rita Bernardelli (ProRett Foundation) and Lucia Monaco (Telethon Foundation) for their support in organizing the Rett survey and field study. We also thank Bernard Munos, Enrico Zaninotto, two anonymous referees and participants at the Sunbelt XXX meeting in Riva del Garda for helpful comments.

“To create consists precisely in not making useless combinations and in making those which are useful and which are only a small minority [...] not that I mean as sufficing for invention the bringing together of objects as disparate as possible; most combinations so formed would be entirely sterile. But certain among them, very rare, are the most fruitful of all.”

(Poincaré, 1921, p. 386)

## 1. Introduction

The evolution of scientific, artistic and economic domains largely depends on the creative *élan* of knowledge communities to attract new talents and to innovate. The centrality of creativity as the engine of social and scientific progress is witnessed by the plethora of research efforts aimed at unveiling its origin and nature. Although such contributions are widespread in various disciplines, it is still possible to identify three main streams of research.

The first approach focuses on the personal or intra-psychic features of individuals revealing a peculiar propensity toward creative thinking and problem-solving (Larkin et al. 1980; Simonton 1988, 2000; Hennessey and Amabile, 2010). The fascinating, original and complex personality of great geniuses, which has often placed them outside so-called normal science and society, is at the origin of this interest.

Beginning in the 1960s, a growing group of scholars has been studying the social factors of creativity, through reappraisal of social contexts and social relationships in stimulating or repressing individual creativity (Kuhn 1962; Merton 1973; Latour and Woolgar 1979; Amabile 1983; Csikszentmihalyi 1990; White 1993). Stretching the social perspective, the concept of creativity has been reformulated in terms of a collective activity which derives from the interaction of two or more individual minds (Kurzberg and Amabile 2000; Shalley and Perry-Smith 2008). This view seems better to suit the recent evolution of natural sciences, in which major advances derive from the creative activity of scientific teams, as shown by the great spread of co-authorships and joint research projects (Newman 2001; Guimerà et al. 2005; Barabási 2005; Wuchty et al. 2007; Jones 2008). Although several studies have focused on individual creativity in social networks (Burt 2004; Uzzi and Spiro 2005; Fleming et al. 2007) and the importance of the network itself as the locus of innovation (Powell et al. 1996; Lane et al. 1996), the collective dynamics of creative problem-solving are still largely under-explored.

A third stream of literature stems from evolutionary biology (Fontana 2001). According to this approach, novelty emerges from the co-evolution of multiple networks (Padgett and Powell 2011). Once a new concept is created, the constraints and opportunities of interaction in multiple social and

semantic networks determine its spread and the possibility of changing the structure of those networks. Creativity is a property of a network of networks.

First attempts to combine the identification and spread of new concepts in a single model of scientific discovery were based on a two-step process: first, some creative scientists originate new ideas, then those ideas spread through a social network as epidemics (Goffman 1966; Valente 1995; Bettencourt et al. 2006). However, this approach does not consider the combinatorial nature of knowledge creation, the co-evolution of ideas and social contacts, and the need for modules and niches in multiple networks to ensure the co-existence of several ideas for the production of further knowledge (Lambiotte and Panzarasa, 2009).

In this paper, we treat creativity as a problem-solving activity characterized by novelty, unconventionality, persistence, and difficulty in problem formulation (Newell et al. 1959). To study creativity, we focus on rare diseases as scientific specialties where problems are *novel* (most the research started after the genomic revolution in the 1990s), *unconventional* (diseases are frequently misdiagnosed), *persistent* (more than a decade is required, from problem identification to a cure) and *ill-defined* (causes and symptoms of rare diseases are still largely unknown). We focus on the process of problem-solving which can be described as a search through a maze (Simon, 1969), requiring the activation and teamwork of scientists (the emergence of communities and team assembly) as well as the identification, combination and retention of useful concepts (abstract and general ideas inferred from specific instances). The more difficult and novel the disease, the greater the amount of trial-and-error required to find a solution by recombining individual capabilities and concepts. We hypothesize that new combinations of scientists and concepts which represent progress toward the goal of finding a cure are retained as stable modules in socio-semantic networks; all others are dissolved. To test our research hypothesis, we perform an in-depth case study and develop a theoretical framework for the evolution of scientific creativity.

The paper is structured as follows. In Section 2, we locate our research in the literature on creativity. Section 3 describes our inductive approach and research methodology: from an in-depth case study analysis of a rare disease (Section 4) to a model of the co-evolution of social and semantic networks (Section 5). The last section is devoted to a discussion and summary of our main findings.

## 2. Knowledge Creation and Diffusion in Scientific Networks

Creativity is a complex phenomenon which has been studied from several different perspectives. The socio-economic literature notes that creative activity is accomplished by people who occupy a special position in the social network or the economic system at a certain moment in time (Schumpeter, 1934). However, there is no consensus on exactly where creative people can be found: at the periphery of the community, at the center, or in a brokerage position.

In his analysis of the evolution of science, Kuhn argues that they must be peripheral, to keep some distance from mainstream knowledge and social relations and to gain a different view of the field: “*Any new interpretation of nature, whether a discovery or a theory, emerges first in the mind of one or a few individuals. It is they who first learn to see science and the world differently, and their ability to make the transition is facilitated by two circumstances that are not common to most other members of their profession. Invariably their attention has been intensely concentrated upon the crisis-provoking problems; usually in addition, they are men so young or so new to the crisis-ridden field that practice has committed them less deeply than most of their contemporaries to the world view and rules determined by the old paradigm.*” (1962, p. 143). As novices, they may offer the field their fresh and divergent thinking. The kind of creativity underlying such discoveries displays a generative nature, as it implies the introduction of new concepts and theories on which new scientific achievements become possible (*generative creativity*).

Conversely, as brokers, individuals are at the crossroads of knowledge circulating within different social circles. In this position, they have the best opportunity to produce creative ideas. In this perspective, creativity has typically a combinatorial nature, as it is defined as a link among concepts and perspectives which developed apart, and the creative person is seen as a bridge between communities which were not originally connected (*combinatorial creativity*). Since the work of Schumpeter (1934), several contributions may be ascribed to this view (Brass 1995; Hargadon and Sutton 1997; Weitzman 1998; Perry-Smith and Shalley 2003; Burt 2004; Fleming et al. 2007; Fleming and Waguespack 2007; Cattani and Ferriani 2008). Conversely, cohesive groups may lead to fewer new combinations but can improve information circulation as well as the diffusion, sharing and generalization of ideas (Gould 1991; Watts 2002) through the *replication* of new combinations of concepts (*replicative creativity*).

Both from an individual viewpoint and for the community of scientists as a whole, the benefits of a structural position (novice, broker, leader) and a certain type of creativity (generative, combinatorial, replicative) may change over time. Following Shalley and Perry-Smith (2008), we



distinguish three phases: (1) problem identification and formulation, (2) conceptual combination and (3) conceptual expansion.

In the first phase (*problem identification and formulation*) information is insufficient and unordered (Goffman and Harmon 1971; Chen et al., 2009). There are multiple ways of framing the problem, and it is essential to identify its main elements. Novel ideas prosper and the selection forces which normally act on them are still weak (March, 2007). In this phase creativity is mainly generative. The second phase (*conceptual combination*) is characterized by the creation of relationships between previously separate concepts across multiple sources, categories and knowledge domains. Creativity here is combinatorial. In the last phase (*conceptual expansion*), new ideas are shared and generalized by means of various techniques such as analogies, metaphors and remote associations. Replicative creativity is prominent at this point.

Apart from a few notable exceptions which reveal the emergence of novelty at various levels of social interaction, few papers try to combine complex phenomena like creativity along more than one dimension, by looking in particular at both the scientific community and the related semantic space (Padgett and Ansell 1993; Orsenigo et al. 2001; Taramasco et al. 2010; Roth and Cointet 2010). This paper aims to define a link between the dynamics of creative thought and the social dynamics of the community. Different streams of literature have devoted attention to either one or the other of these dimensions, implicitly declaring an order of relevance. Typically, in the social sciences the focus was on social mechanisms, while in the human sciences more attention was addressed to creative thought. Against this background, we try to restore some of the complexity of the creativity phenomenon and its multifaceted and inevitably intertwined dimensions by considering a co-evolution hypothesis.

To operationalize our approach, we consider three types of change in both semantic and social networks (see Table 1):

- *Generative*: the entry of new nodes (new concepts/authors);
- *Combinatorial*: the creation of new links (new scientific collaborations or new concept associations);
- *Replicative*: the replication of connections (multiple collaborations among the same scientists and repeated use of concept associations).

[INSERT TABLE 1 ABOUT HERE]

Here, we first perform a case study of a single specialty to validate our definition of creativity. More precisely, we aim at testing whether the most creative scientists are those who first introduced

new concepts or new associations of concepts. Next, we develop a theoretical framework to study the co-evolution of semantic and social networks. Specifically, we analyze the relationships between the creation of nodes, links, and link replication in both semantic and social networks.

### **3. Empirical protocol, data and methods**

#### ***3.1 Research strategy overview***

The goal of this paper is to define, measure and model creativity in scientific networks. In the first part of our work (Section 4), we conduct an in-depth analysis of a single scientific community. In the second part (Section 5), we develop and test a stochastic model of network evolution.

The case study is designed: (1) to identify the conceptual cores and field boundaries of an epistemic network (Laumann et al. 1983); (2) to explore the possibility of measuring creativity by means of peer evaluation, bibliometric and network indicators; (3) to test whether the most creative scientists, identified by measures at point (2), have introduced new combinations of concepts in the evolution of a rare disease specialty toward a cure. Throughout the case study, we combine both qualitative and quantitative methods, such as archival and bibliometric analysis, to benefit from their complementarities (Denzin and Lincoln 1994) and to achieve reciprocal internal validation of results (Yin 2004).

Our modeling effort builds upon a tradition of stochastic null-models of science dating back to Simon (1955). The model is based on three parameters: the entry of new nodes, the creation of new links, and the replication of existing links (Guimerà et al. 2005). First, we simulate the model for the social network of co-authorships and the semantic network of concept co-occurrence in scientific papers. Then we estimate the maximum likelihood value of the three parameters in several rare diseases. Lastly, we check for any relationship between the structural parameters of the model for the semantic and social networks across different domains. The presence of a clear-cut relationship among them implies that the two evolutionary processes of the semantic and social networks are interdependent.

#### ***3.2 Research setting***

In order to be theoretically and methodologically coherent with our research questions, a suitable empirical field of study should display creative problem-solving of different kinds

(generative, combinatorial, replicative). It should also provide a moderately sizable and well-defined research topic. The first focus is on one rare disease occurring almost exclusively in girls, called the Rett Syndrome (RTT). Next, we test our model on more diseases: Rett, Noonan and Horner Syndromes, Mesothelioma, Paroxysmal Nocturnal Hemoglobinuria, and Adenosine Deaminase deficiency. We tested the adequacy of RTT, a rare disease, as a research setting so that it could be empirically defined as a creative and closed field. In the following subsections, we describe how this was done, reporting results in Section 4. The same method was applied to identifying the other five rare diseases in Section 5.

### ***3.3. Empirical closure***

According to Csikszentmihalyi (1990), creativity requires dynamic interactions among three subsystems: the *individual* creator, the knowledge *domain* (i.e., the Kuhnian paradigm), and the *field*, which consists of those persons who work in the same domain, and thus have their creativity governed by the same domain-specific guidelines. As a field for our study, we chose a *specialty*, defined as “a group of researchers and practitioners who have similar training, attend the same conferences, read and cite the same body of literature” (Fuchs, 1993). This choice was made for two reasons. Conceptually, it is correct to segregate science into subfields displaying cohesive dynamics of both knowledge concepts and scientists working within them. Scientists do select their readings and relations according to their interests, and thus identify a specific reference community and their shared knowledge basis (Börner et al. 2004). In practice, this allowed us to solve the methodological problem of defining the boundary of the epistemic network, by selecting a field which is computationally tractable, but also socially and thematically closed.

To trace the boundary of the RTT specialty, we examined whether any convergence of the field (i.e., important topics and reference people) existed, according to experts’ assessment of key scientists and papers collected through questionnaires, lists of topics addressed in specialized conferences, scientific reviews, and numerous descriptions of the field which are available to the general public through the internet and other informational sources. We also mined all papers in PubMed mentioning the term “Rett Syndrome” in the abstract and cross-checked the two lists.

### 3.4. Data availability

Another reason for choosing rare diseases, and RTT in particular, is due to their recent history, which can be traced both on the web and in databases such as PubMed since their early beginnings. Eighty per cent of rare diseases are of genetic origin. Since the Orphan Drug Act was passed in 1983 and the Human Genome Project started in 1990, most research in this field has been carried out in the last 25 years. To sum up, the choice to study rare diseases appeared to be suitable: they are new specialties of tractable dimensions.

### 3.5 Data collection

Data for case-study analysis on RTT was collected in four steps.

First, we decided to conduct research on existing scientific and informational literature on the evolution of RTT knowledge, in order to extend our understanding of the dynamics of the field, without being *a priori* guided by the beliefs of RTT experts. We referred to both specialized and non-technical materials published on the web (especially parents' associations websites) and in scientific journals. We also attended the 2009 Meeting of the European Working Group on Rett Syndrome<sup>2</sup>, where we also recorded the sessions and checked for the persistence of key concepts/scientists in the discourse (number of times they were mentioned in presentations).

Second, we contacted a set of RTT experts and proposed a questionnaire to scientists working in the field (see Appendix 1). Our goal was to validate the existence of a social community (*field*) and common semantic space (*domain*) corresponding to the specialty analyzed. In this way, we aimed at ascertaining the existence of a convergence on what they consider their reference context, in terms of both knowledge and people. In addition, we used these data to build our understanding on an endogenously and empirically defined notion of creativity. As regards the identification of interviewees, a primary list of contacts resulted from the combination of the contacts of two Research Foundations on Rett Syndrome and contact with 78 RTT authors listed in PubMed. The list was integrated with participants in one of the main international conferences in the field (63 names). The questionnaire was structured with open questions, and was sent to contacts by e-mail. We asked the RTT experts to rank at least five authors they considered the most important in the field, and the five most creative<sup>3</sup>. We then asked them to cite at least five key scientific publications in the field and the most important research topics. Our response rate was 11%, including 5 of the

---

<sup>2</sup> [www.rettmeeting.org/](http://www.rettmeeting.org/)

<sup>3</sup> We did not provide our interviewees with a definition either of relevance or of creativity, as the aim of the inquiry was to find a definition embedded in the community.

top 30 RTT scientists. Although the results are not statistically significant, given the small size of the community and the low response rate, all responders were experts with publications in this field.

Third, we extracted some classical bibliometric indicators from Scopus (Hirsch's *h*-index) and the ISI Web of Knowledge (the sum of the journal impact factor of authors' publications) and compared them with data on creative scientists and contributions collected through the questionnaires.

Fourth, to analyze semantic and social networks, we developed a thesaurus of rare diseases which could be adapted to the great fragmentation of knowledge in the field. We then extracted all the publications on rare diseases in PubMed. We analyzed the dynamics of semantic and social networks and identified the connections between the two, tracing the relationship between authors and their creative concepts. We developed a tool similar to ConceptLink (White et al. 2004) to generate concept maps with the Unified Medical Language System (UMLS) co-occurrence database of MeSH descriptors. Thus, one node in our semantic network is a concept (MeSH term) and links among concepts are computed according to the number of co-occurrences of a pair of concepts in PubMed abstracts. In a similar fashion, we built a co-authorship network based on the number of joint publications of biomedical scientists, and computed three simple measures of author centrality: strength, degree, and *k*-core. Centrality indicators were selected to check whether authors with several joint publications (strength), more co-authors (degree) and members of the inner core of the network (*k*-core members) were the most creative scientists.

## **4. The Rett Syndrome**

### ***4.1 Case description***

In the field of biomedical sciences, rare diseases are emerging research specialties. According to the European definition, a rare disease is a serious or lethal illness affecting less than one person in every 2,000 individuals (75 cases out of 100,000 individuals according to the U.S. definition). Recent estimates show that not less than 30 million Europeans suffer from rare diseases. Nevertheless, the field is far from maturity. In fact, despite rarity, five new rare diseases are discovered every week and, for 8,284 diseases out of 9,471 (87 per cent), there is no information available in public data sources and no ongoing research activity ("orphan diseases"). After the Orphan Drug Act was passed in the United States (1983), research in some rare disease areas has intensified. However, most rare diseases are still specialties in which creativity and novelty are pivotal, for both identification of the disease, and any progression toward a cure.

Our case study focuses on the Rett syndrome (RTT), a genetic neurodevelopmental disorder which appears in infancy and predominantly affects girls. RTT patients are normal at birth and during early development but, after the sixth month of age, display postnatal deceleration of head growth, psychomotor regression, gait dysfunction, and stereotypic movements such as the so-called “hand-washing”. RTT is the second cause of mental retardation in girls throughout the world, with an estimated average incidence of one case every 22,800 girls between 2 and 8 years old. Unfortunately, there is no treatment available for RTT. Several steps have been taken toward identifying treatment<sup>4</sup>: they are based, first, on a clear-cut definition of the disease and, second, on deep understanding of its genetic mechanisms. At present, knowledge of RTT and its genetic causes is still unstable, as it is constantly being revised and refined<sup>5</sup>.

On the question of definition, there is a substantial consensus regarding the descriptive traits of the disease. However, neither the International Classification of Diseases nor the Diagnostic and Statistical Manual of Mental Disorders appropriately classify RTT in line with recent evidence. RTT is not an autistic disorder, although it is categorized among them. As regards identification of genetic causes, 95% of classic RTT cases reveal a mutated gene in the X chromosome, which provides instructions for the over- or under-production of a protein (MeCP2) which is critical for normal brain development. Conversely, only 20-40% of girls affected by a variant form also show a MeCP2 mutation. As a result, other genes, such as CDKL5, have been highlighted as responsible for a set of such RTT syndrome manifestations. Nevertheless, in general, such mutations are not one-to-one associated with RTT, but may also occur in other diseases.

While the causes of the syndrome are challenging researchers, progressive ability to identify and diagnose RTT has enriched the picture of cases<sup>6</sup> and has also added complexity to their potential understanding. For example, not only girls are affected by RTT, as some cases of male phenotypes have also been discovered. Overall, we may state that knowledge of RTT is at a stage in which there are some descriptions of empirical facts, such as clinical evidence, sometimes associated with behavioral traits, and that these facts are sometimes ordered into clearly defined new diseases over

---

<sup>4</sup> The first phase consists of identifying a set of typical characters of the disease (1a) and its association with the genetic disorder which is responsible for its onset (1b). As a second stage, precise understanding of the molecular mechanisms underlying the disease is required (2), which recognizes molecules and therapeutic strategies to be tested in vitro (3), and later in laboratory animals (4). After these four steps (pre-clinical phases), it is possible to test the treatment on humans and results will show whether any treatment for the disease has been found or not.

<sup>5</sup> See, as an example of such a process of redefinition and specification, changes in the topics list concerning the European Working Group on Rett Syndrome First and Second Conferences, which took place respectively in 2007 and 2009: 2007 Topics - The Molecular Cause of Rett Syndrome, MeCP2 Target Genes, Respiration Control and Seizures, Neuronal Plasticity, Future Approaches; 2009 Topics - Basic Molecular Mechanism of MeCP2, Circuit Defects in Rett Syndrome, Behavioral Deficits in Mice Lacking a Functional MeCP2, Modifier Genes and Other Genes Involved in Rett Syndrome, Therapeutic Approaches, Molecular Genetics.

<sup>6</sup> This information is now more easily available to researchers and physicians, as a database on MECP2 mutations (RettBase) and a repository of clinical information (InterRett) are being collected.

time. However, knowledge is far from consolidated. Conversely, continuous growth supported by creative discoveries is acknowledged.

#### ***4.2 Is Rett Syndrome a specialty?***

By the end of 2009, 4,498 researchers had contributed to research on RTT by producing 1,653 scientific articles and 2,243 distinct concepts (MeSH terms). Table 2 lists some key statistics of the top scientists in the RTT research domain. They are the most frequently cited scientists by our responders, ordered by year of entry in the RTT community (last column of the table). We highlighted those who made the most significant discoveries about RTT. Questionnaire results, bibliometric indices and network statistics are reported. With reference to questionnaire results, for each attribute we show both ranking position and a score proportional to the number of times a scientist was mentioned and the number of scientists the single responder cited in the answer (columns 1-4). Columns 5-6 reflect the frequency with which a paper written by a scientist was mentioned among the key papers. We also provide an average of the first three variables (columns 7-8). With reference to bibliometric analysis, we report overall production (number of papers) and the Hirsch's *h*-index of each author, as well as the number of papers on RTT and the sum of the Impact Factor of RTT publications. As network statistics, we report strength, degree, *k*-core and year of entry into the network.

The set of key scientists, as recognized by peers in questionnaire results, is restricted to the group of authors listed in Table 2. Adrian Bird and Huda Zoghbi are unanimously recognized as the most important scientists in this field; Andreas Rett (the Austrian scientist who first identified the syndrome) ranks fifth. Higher dispersion is displayed with regard to publications; however, a core of four most important works can be clearly identified. All respondents except one cited Guy et al. (2007) among the five most important readings in the field. Only two did not mention MeCP2 as a key concept in the field. On key concepts for research, both questionnaires and analysis of conference topics showed that research is focused on the study of MeCP2, the most probable main cause of RTT. Both informative and specialized literature accounts for the evolution of studies on RTT, citing the same scientists as keys to the field and the same open challenges for research.

Overall, these results support the hypothesis that RTT is a specialty, with a definite reference community (*field*) and knowledge base (*domain*). Scientists display high convergence of perceptions on people and topics (community closure). People in the epistemic community read the same papers, share similar research interests and relate to the same colleagues.

[INSERT TABLE 2 ABOUT HERE]

#### *4.3 How can we measure creativity?*

The lists of the most creative scientists and of the most important scholars identified by community members do not perfectly match (see columns 1-4 in Table 2). In addition, interviewees never report the same names and/or rankings for the two questions. This shows that the difference between the two attributes has been carefully weighted by responders. Comparing the ranking of the top ten most creative and most important scholars, we note that creativity is associated with more recent contributors (although their papers are not considered to be references in the field); as regards importance responders recall the “founding fathers” of the field in question: Rett and Hagberg.

The most creative scientist<sup>7</sup>, Adrian Bird, has recently shown the reversibility of RTT, and thus the possibility of finding therapy for patients (Guy et al., 2007). Next in the ranking, Huda Zoghby has discovered that the mutation of the MeCP2 gene is one of the major causes of RTT Syndrome. Michael Greenberg has directed attention to other genes which may be related to MeCP2 and be the indirect cause of the disorder. Rudolph Jaenisch has contributed to the advances in knowledge of RTT. By introducing mouse models and using transgenic animal experiments as suitable methods for gene research, he has shown that therapeutic cloning can correct genetic defects in mice. He leads a group which has patented a candidate treatment (IGF-1) and has also recently started a clinical trial<sup>8</sup>.

All these contributions lie on the path to definite understanding of the disorder and aim at finding treatment. On this point, it seems that works which approach a cure (i.e., provide an answer to the key research question) are clearly assessed as creative by the scientific community and are very well considered by the community of patient families and donors which socially and financially supports such research. It also seems that creative scientists contribute new concepts or establish new associations between key concepts to find a cure for RTT. Bird introduced the new concept MeCP2 (generative creativity). MeCP2 was then linked to RTT by Zoghby (combinatorial creativity). Authors who have introduced new concepts or new combinations of concepts are recognized as creative by the scientific community. These two dimensions build up our endogenous definition of creativity, which also reflects its generative and combinatorial nature, as pointed out in the literature.

---

<sup>7</sup> Cfr. Table 2 highlighted scientists.

<sup>8</sup> <http://www.disabled-world.com/medical/clinical-trials/rett-trial.php>.



Although we found a correspondence between descriptions of RTT and experts' evaluation of creativity, it is not easy to find a good measure of creativity from standard bibliometric indicators such number of papers, Impact Factor (IF), *h*-index and centrality of authors in the community. This is particularly challenging in the case of young scientists. Centrality indicators in the network of co-authorships, such as strength or degree, are only loosely related to creativity ranking provided by peers. For instance, the scientist with the highest strength score was not even mentioned by our responders as being a creative person. The number of years of contributions to the field do not appear to be significant in relation to creativity, whereas IF scores (especially when divided by production time) and the *h*-index can replicate the very top positions (like Bird and Zoghbi), but appear to be highly unsuitable to capture the phenomenon as a whole<sup>9</sup>. As a result, there is no direct mapping of creativity as has been “declared” by experts with a single bibliometric indicator (van den Beemt and van Raan 1995; Rinia et al. 1998).

Another reason why classical metrics are not effective in signaling creativity may also lie in the evolutionary path which it has taken within the field. Both the deployment of creativity and the involvement of authors on RTT have taken place over time in the form of waves, which have given new impulse to the study of RTT and have enhanced knowledge.

The triangulation of narrative, survey results (and analysis of their correspondence, described previously) and the average IF per year allows us to claim that there are different generations of creativity within the evolutionary path of RTT knowledge, in which there is one most creative author in each phase, a person who is linked to the introduction to a new concept or a new combination of concepts. If we trace the temporal evolution of the network, we note that, by introducing new (combinations of) concepts, such authors become the most prominent authors of their cohort.

One last consideration must be added with regard to the results of the questionnaires. We noted that the answers to the four simple questions are affected by what could be called a “recency effect” with reference to the behavioral literature. In 90% of cases, the subjects cited scientists, publications and research topics which have appeared in the field in the last decade, whereas the creative contribution, for example, of the scholar who identified Rett Syndrome as a rare disease is mentioned only once<sup>10</sup>. This supports the hypothesis that there is a thematic and temporal range which scientists implicitly consider to be relevant, which makes them discount creative or

---

<sup>9</sup> Adrian Bird, who was identified as the most creative and important, has the highest IF, but other indicators do not correspondingly highlight such a prominent position in the network. This weak correspondence is lost for other authors who are ranked lower, such as Michael Greenberg or Rudolph Jaenisch. Young researchers, e.g. Monica Justice, were mentioned by peers for their creativity, but are not present in the network yet. Conversely, scientists Wade and Bienvenu have a significant position in the community, as reflected by bibliometric and network indicators, but this role is not elicited by the community when directly asked.

<sup>10</sup> The most cited paper of Adrian Rett on MeCP2 is never mentioned among the top readings in the RTT field.

influential but “old” contributions, although they were central in the field taken as a whole (Börner et al., 2004). Old contributions and contributors are present in cumulative networks built on data drawn from publication datasets. The understanding of time windows is very important from a methodological point of view, for calibrating the representation of such networks in a dynamic way, so that what is forgotten or remains in the background can be deleted from the picture.

#### ***4.4 Who are the most creative scientists?***

We distinguish two types of creativity: generative and combinatorial. Generative creativity is defined as the introduction of new concepts. Combinatorial creativity is the creation of new relationships among concepts. To measure creativity, we use a paper as the basic unit of analysis. A paper may be seen as a set of authors ( $a$ ) and concepts ( $c$ ):

$$p(c_1, \dots, c_i, \dots, c_n / a_1, \dots, a_j, \dots, a_m)$$

The collection of papers about RTT defines a socio-semantic or epistemic network. A paper is a hyperlink in the co-authorship network (Tamarasco et al. 2010) and, at the same time, a hyperlink in the semantic network of concept co-occurrences. More in general, it defines a set of relationships in the socio-semantic network of concepts and authors. This approach may be generalized to the case in which more than two multiple categories are considered (such as citations and research methods). To be published, a paper must contain some degree of novelty (generative, combinatorial, or both). By analogy, we can identify two changes in the network of co-authorships:

1. A paper contains a new author: a new node is added to the co-authorship network;
2. A paper contains a new combination of authors: a new link is added to the co-authorship network.

In the evolution of the socio-semantic network, new nodes (authors/concepts) enter and new links are formed. New authors can either be already active scientists entering a given specialty from other scientific communities, or new researchers. Thus, we must further distinguish new nodes in the scientific community from new nodes in the specialty. Let us identify new actors in a given specialty  $s$  by  $n_s(a)$ , of which  $n(a)$  have never published before, and new scientific collaborations (co-authorships) by  $n(a^2)$ . Similarly, generative creativity is the entry of new concepts  $n(c)$ , whereas combinatorial creativity has to do with the creation of new links among previously unrelated concepts  $n(c^2)$ . As before, we can distinguish locally new combinations  $n_s(c)$  in a given specialty from globally new concepts and concept associations in science. The degree of novelty of a paper may thus be measured as:

$$n_s(p) = \{t(c), t_s(c); t(c^2) \mid t(a), t_s(a); t(a^2)\} \quad \text{for all (the combinations of) } a \text{ and } c \text{ in } p$$

where:

- $t(c)$  is the number of previous occurrences of concept  $c$  in the literature (e.g. PubMed);
- $t_s(c)$  is the number of previous occurrences of concept  $c$  in a given specialty (e.g. RTT);
- $t(c^2)$  is the number of previous co-occurrences of two concepts;
- $t(a)$  is the number of previous papers of author  $a$ ;
- $t_s(a)$  is the number of previous occurrences of author  $a$  in a given specialty;
- $t(a^2)$  is the number of times two authors have worked together;

For instance, a hypothetical paper with novelty  $n_s(p) = \{(9, 6), (7,2); 0 \mid 10, 20; 4\}$  has two concepts and two authors. The first concept has been used 9 times (6 in specialty  $s$ ), the second 7 times (of which 2 in specialty  $s$ ), but this is the first time the two concepts are used together. The authors have already published 10 and 20 papers respectively, of which 4 are co-authored. The creativity of a given author can now be measured by looking at the  $n_s(p)$  statistics of its papers.

[INSERT FIGURE 1 ABOUT HERE]

In PubMed we found 1,653 papers on RTT, with about 4,500 authors and 2,300 concepts. On average, each paper has 4.73 authors and 12.23 concepts. The most central specific concepts in the RTT network are Rett Syndrome and MeCP2, which were discovered by Rett and Bird, respectively. The first link between the two concepts was established by Zoghbi. Jaenisch demonstrated the association between Rett and MeCP2 in a mouse model (first Rett-Mice-MeCP2 combination). More recently, Bird has shown the reversibility of the pathology in mice (first Rett-Therapy combination, albeit limited to transgenic mice). The distributions of the share of new combinations of authors/concepts per paper overlap almost perfectly (Figure 1). This result indicates a positive relationship between the openness of scientific teams to new collaborations and combinatorial creativity at the community level.

[INSERT FIGURES 2-3 ABOUT HERE]

The evolution of the RTT network tells us that the centrality of a scientist in a community is associated with the ability to introduce new useful concepts or concept associations. However, we find a negative relationship between seniority and the degree of novelty of the scientific concepts in question (Figure 2). Conversely, experience has no effect on combinatorial creativity (novelty of

concept combinations). Also, as Figure 3 shows, a positive relationship exists between the authors' centrality and the number of publications which used the new (associations of) concepts they introduced. First, to become central, researchers must introduce new (combinations of) concepts. Second, the centrality of creative scientists in the co-authorship network is driven by the replication of concepts they have brought to the field.

In the first stage of emergence of RTT as a specialty (problem identification), the rate of entry of new authors, mainly from other areas of research, was high. The first phase was completed with the introduction of new concepts (e.g., Rett and MeCP2), based on which the field has been established. The introduction of new concepts spurred the combinatorial creativity of the community and gave rise to the emergence of a giant component through new scientific collaborations (concept combination and extension).

## 5. A stochastic model of scientific creativity

### 5.1. A baseline model for co-evolution of networks

The structure of scientific activity has been intensively studied over the last fifty years. On one hand, several empirical regularities of the networks of scientific collaborations have been identified, on the other, simulative and theoretical models able to replicate the main features of scientific networks have been developed. Building on a tradition of science modeling dating back to Simon (1955) we generalize previous stochastic growth models to describe the dynamics of socio-semantic networks. Let us first consider a production function of the form:

$$\dot{C}=f(C,A)$$

where  $C$  is the stock of knowledge,  $\dot{C}$  the amount of new knowledge produced at time  $t$ , and  $A$  the number of active researchers<sup>11</sup>. Since one of the principal constraints to publication in science is that no two papers which contain the same knowledge may be published, new papers must contain new knowledge. Hence, science as an institution may be described by the production of papers, each proposing a new “quantum of knowledge”:

$$P(\dot{C})=f(P(C),A)$$

where  $P(C)$  is the stock of all papers available and  $P(\dot{C})$  new papers produced at time  $t$ . Gilbert (1997) shows that it is possible to generate many of the quantitative features of the present structure of science by looking at scientific activity as a system in which scientific papers generate further

---

<sup>11</sup> More in general  $L$  denotes all rivalrous production inputs, such as labour, and  $C$  all non-rivalrous ones, such as ideas.

papers, with authors playing a necessary but incidental role. Börner et al. (2004) also rely on an author-paper production function to explain the evolution of scientific networks of co-authorship and citations. In our model, a paper is a combinatorial set of concepts and authors

$$p(c_1, \dots, c_i, \dots, c_n | a_1, \dots, a_j, \dots, a_m)$$

We build on the growth model originally presented by Simon (1955) in the version modified by Guimerà et al. (2005) to explain team assembly in science. At any time step  $t$ , a new paper is produced. In our stochastic model, we select authors and concepts at random based on the following probabilistic rules. Some concepts and authors in the paper may be globally new in science or locally new for a given specialty. We start the simulation with an endless pool of new authors and concepts. First we select  $m$  authors.<sup>12</sup> A new author in a given field of research (topic) is called a *novice*. Novices become *experts* after their first publication. Each member of the team has probability  $p_1^a$  of being a novice and probability  $1-p_1^a$  of being an expert. If an author is drawn from the experts' pool and there is already another expert in the team, with probability  $p_2^a$ , a new combination of experts is established otherwise with probability  $1-p_2^a$ , the new author is randomly chosen among the set of collaborators of a randomly selected team member.<sup>13</sup> The probability of an expert being selected is proportional to the number of papers that expert has written. The same procedure is applied to the concept set. We select  $n$  concepts. We add a new concept (generative creativity) with probability  $p_1^c$  whereas with probability  $1-p_1^c$  a concept which has been already used is added to the paper. In the second case, with probability  $p_2^c$ , a new concept association is formed (combinatorial creativity); otherwise an already existing combination of concepts is replicated with probability  $1-p_2^c$ . The likelihood of selecting an existing concept is proportional to the number of times it has already been used. We apply the same rules for all authors and concepts of a paper and for all papers. As in the model of Guimerà et al. (2005) this generalized version replicates some of the main topological properties of the co-authorship network: the presence of a giant component encompassing about half the authors, a small-world structure, and a power law connectivity distribution with and exponential cut-off (Lodka, 1926; Girman and Newman 2002; Newman, 2004). In our stochastic framework we can also test the relationship between the four main stochastic parameters of the model:  $p_1^a$ , new authors;  $p_2^a$ , new collaborations;  $p_1^c$ , generative creativity;  $p_2^c$ , combinatorial creativity.

---

<sup>12</sup> In our model the number of authors and concepts per paper are positive random numbers with mean  $n$  and  $m$  respectively.

<sup>13</sup> The probability  $p_1 = 1-p$  and  $p_2 = 1-q$  in the model of Guimerà et al. (2005). We have modified that model to accommodate teams of different sizes. As in the original model, agents who remain inactive for longer than  $T$  time steps are removed from the network. However, our results do not depend on the specific value of  $T$  which is usually set to the maximum level, since we are analyzing new fields of research in which the vast majority of authors are still active.

## 5.2. Simulations and model validation

Based on our analysis of RTT, we expect that a relationship is in place between the evolution of the network of scientific collaborations and creativity, as measured by conceptual generation and combination. In particular, we state that, in a given community, new collaborations and combinatorial creativity are positively related.

(H1).  $p_2^g = p_2^f$ . The probability of new co-authorships is the same as the probability of combinatorial creativity. It should be noted that, unlike the case illustrated in Tamarasco et al. (2010), this relationship holds in probability for a set of related papers and not for a single paper.

The second relationship is more complex. Figure 2 shows that new authors have a higher probability of introducing new concepts. However, only a few of them will be replicated by others, thus contributing to the solution of the problem and the centrality of the creative scientists (Figure 3). Authors like Bird are recruited in the RTT specialty as the new concept they introduce contributes to the RTT problem-solving activity. These events occur rarely, but they are extremely important. Instead, most of the novices are recruited to help experts to replicate their findings.

(H2). In a scientific community  $p_1^g$  and  $p_1^f$  are inversely related. The larger is the participation of novices in scientific production, the smaller is generative creativity. However, at individual level, newcomers have a higher chance of working on new concepts and theories. A high turnover of scientists prevents the emergence of the giant component of experts (community closure) which is crucial for the generation of new concepts.

To test our predictions, we selected six rare disease specialties of different sizes and at different stages of evolution, from problem identification to the development of an effective treatment. The stage of development of a research field toward a cure can be identified by looking at the number of active clinical trials and available treatments. Size can be measured by looking at the prevalence rate (i.e., the total number of cases divided by the number of individuals in the population). For all the rare diseases we considered, except PNH, there is no new treatment in Europe or the United States. Horner and Noonan are in a stage of early development (no active clinical trials). ADA and Rett are at an intermediate stage: there is growing consensus about the causes of the pathology and potential intervention strategies, but only a few trials are ongoing. The fields of mesothelioma and

PNH are more mature (several trials are ongoing, and a new drug was launched in 2007 to treat PNH). At each stage of evolution, we selected one relatively large specialty (Noonan, Rett, Mesothelioma) and one small (Horner, ADA, PNH) (see Table 3).

The number of publications ranges from 827 (ADA) to 9,625 (Mesothelioma); the average number of authors and concepts per paper is relatively stable - around 4 and 11 respectively. The topological properties of the co-authorship networks are widely diversified: size in terms of number of papers, the fraction of authors with at least two publications ( $F_R$ ), repeated co-authorships ( $f_R$ ) and the share of authors in the giant component of the network ( $S$ ) range from 5.88 to 28 ( $F_R$ ), 1.36 to 14 ( $f_R$ ) and 2.23 to 59.58 ( $S$ ), respectively.

We simulated our model for each specialty by taking the average numbers of authors and concepts as inputs ( $n$  and  $m$ ) to generate several random networks for different values of  $p_1^a$ ,  $p_2^a$ ,  $p_1^c$  and  $p_2^c$ , ranging from 0 to 1. For each of them, we fitted four statistics of the simulated network to the same statistics for the real-work semantic and social networks (size,  $F_R$ ,  $f_R$ ,  $S$ ). Lastly, we computed the maximum likelihood values of  $p_1^a$ ,  $p_2^a$ ,  $p_1^c$  and  $p_2^c$ . Results are listed in Table 3.

[INSERT TABLE 3 ABOUT HERE]

First, the probability of new combinations of concepts and authors are the same across all the specialties considered. Therefore, combinatorial creativity (i.e., new combinations of concepts) and the openness of the team assembly mechanism to new partnerships among incumbent scientists are closely and positively related. Second, the lower the entry rate of new authors, the higher the entry rate of new concepts. The generation of new concepts implies closure of the community, repeated collaborations, and the emergence of a specialty (a giant component composed by an invisible college of authors focusing on a specific discipline). Third, generative creativity is partially independent of combinatorial creativity. Our preliminary results indicate that generative creativity precedes combinatorial creativity, but more work is needed to better understand the dynamic interplay between the two forms of creative enterprise.

All in all, combinatorial creativity in a specialty has some preconditions: (1) a domain must be established around a set of new concepts (generative creativity); (2) a field of research composed of a stable nucleus of scientists must emerge (specialty); (3) scientists must be free to enter into new scientific collaborations to explore promising new concept associations.

## 5 Concluding Discussion

In this paper we aimed at establishing a dynamic relationship between scientific collaborations and creativity. Although we are aware that our results must be further corroborated by testing the model in different disciplines and at different stages of scientific evolution, we believe our findings have several implications.

First, we develop a methodology to identify and measure different kinds of scientific creativity. Many earlier studies compare bibliometric results with the judgments of scholars or experts on the quality of research. Most of them find a reasonable correspondence, although a poor correlation was found between citation indicators and the originality of research proposals in applied research (van den Beemt and van Raan 1995). In our case study of the Rett Syndrome community, we also find that traditional bibliometric and network centrality indicators do not accurately measure creativity. Hence, we introduce some new indicators, based on counting the new (association of) concepts a scholar contributes to a given domain of study, which match the result of our expert survey on creativity better. More work is needed in this direction to control for the recency effect and to investigate the relationship between generative and combinatorial creativity in scientific careers, teams and communities.

Second, we develop a stochastic model of network evolution. It replicates most of the topological properties of socio-semantic networks, including the rate of generative and combinatorial creativity. In addition, our null-model can be used to measure the expected level of creativity across domains. Unbiased statistical tests to compare the creative performances of teams of scientists in different research areas can be developed.

Third, independently of the peculiarities of each domain, we find that new collaborations among experienced scientists enhance combinatorial creativity. On the positive side, there are several reasons why the increase of spontaneous scientific collaborations must be related to combinatorial creativity rather than to generative creativity. On the negative side, as stated in the initial quotation of Poincaré, most efforts to build scientific networks through centralized public incentive schemes are doomed to failure. The two combinatorial processes of team self-assembly and creative discovery are closely interwoven. Owen-Smith et al. (2002) show that, in the United States biomedical innovation system, the combination of new concepts is favored by blurring boundaries between basic and goal-oriented research, disciplines and institutions, whereas European research institutes are more frequently organized hierarchically by scientific field. In this paper, we further corroborate their findings by showing that combinatorial creativity is related to open re-combination of scientific teams. Local closure of a community in space, time and by specialty, and



the mobility and collaborations of scientists across disciplines and intuitional boundaries are two sides of the same coin (Coward and Jonard, 2003; David, 2003).

Fourth, generative creativity is linked to the emergence of a scientific community (a stable giant component of repeated ties in the co-authorship network). Our analysis of rare diseases reveals that the community closure around a set of relatively stable and shared concepts is a fundamental prerequisite for creativity. Too high a rate of scientific turnover prevents identification of important new problems and concepts (Bruckner et al. 1996).

Further research is needed to better understand the dynamic relationship between generative and combinatorial creativity in the evolution of science. Some of the newly generated (combinations of) concepts are selected by scientists and replicated. The centrality of a scientist in a given field is driven by the replication and validation of the new associations of concepts that person contributed to that knowledge domain. In this respect, our study is complementary to that of Bettencourt et al. (2008) regarding the spread of scientific ideas. It would probably be useful to extend our framework to analyse replication dynamics in citation networks and also to examine whether scientists in different disciplines and institutional settings differ in terms of strategies for successful problem-solving: some will leverage social connections to be creative as a consequence of their ability to collaborate and recruit talented scholars, whereas others may focus on combinatorial creativity and collaborate with colleagues to validate and replicate their findings. This relates to the broader problem of distinguishing homophily from contagious diffusion in networks. Lastly we hope that our work may contribute toward sustaining creative problem-solving in rare diseases.

## References

- Amabile, T.M. (1983). *The social psychology of creativity*. Berlin: Springer-Verlag.
- Barabási, A. (2005). Network theory. The emergence of the creative enterprise. *Science*, 308, 639-641.
- van den Beemt, F.C.H.D., van Raan, A.F.J. (1995). Evaluating research proposals. *Nature*, 375, 272.
- Bettencourt, L.M.A., Kaiser D.I., Kaur, J., Castillo-Chàvez, C., Wojick, D.E. (2008). Population modeling of the emergence and development of scientific fields. *Scientometrics*, 75(3), 495-518.
- Börner, K., Maru, J.T., Goldstone, R.L. (2004). The simultaneous evolution of author and paper networks. *The Proceedings of the National Academy of Sciences of the United States of America*, 101, 5266-5273.
- Brass, D.J. (1995). Creativity: It's all in your social network. In C.M. Ford & D.A. Gioia (Eds.), *Creative action in organizations* (pp. 94-99). Thousand Oaks, Ca: Sage.
- Bruckner, E., Ebeling, W., Jiménez Motaño, M.A., Scharnhorst, A. (1996). Nonlinear stochastic effects of substitution: An evolutionary approach. *Journal of Evolutionary Economics*, 6, 1-30.
- Burt, R.S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110, 349-399.
- Cattani, G., & Ferriani, S. (2008). A core/periphery perspective on individual creative performance: Social networks and cinematic achievements in the Hollywood film industry. *Organization science*, 19(6), 824-844.
- Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3(3), 191-209.
- Csikszentmihalyi, M. (1990). The domain of creativity. In M. A. Runco & R. S. Albert (Eds.), *Theories of creativity* (pp. 190-212). Newbury Park, CA: Sage.
- Cowan, R., & Jonard, N. (2003). On the workings of scientific communities. In A. Geuna, A.J. Salter & W.E. Steinmueller (Eds.), *Science and innovation: Rethinking the rationales for funding and governance* (pp. 309-333). Cheltenham, UK: Edward Elgar.
- David, P. (2003). Cooperation, creativity and closure in scientific research networks: Modeling the dynamics of epistemic communities. In J.-P. Touffut (Ed.), *Institutions, innovation and growth: Selected economic papers* (pp. 170-206). Cheltenham, UK: Edward Elgar.
- Denzin, N.K., & Lincoln, Y.S. (1994). *Handbook of qualitative research*. Thousand Oaks: Sage.
- Jones, B.F. (2008). The burden of knowledge and the 'death of the Renaissance man': Is innovation getting harder?. *Review of Economic Studies*, 76(1), 283-317.
- Fleming, L., & Waguespack, D.M. (2007). Brokerage, boundary spanning, and leadership in open innovation communities. *Organization Science*, 18(2), 165-180.
- Fleming, L., Mingo, S., & Chen D. (2007). Collaborative brokerage, generative creativity and creative success. *Administrative Science Quarterly*, 52, 443-475.
- Fontana, W. (2001). Novelty in evolution. *Bioevolutionary concepts for NASA*, BEACON.

- Galenson, D., & Weinberg, B. (2001). Creating modern art: The changing careers of painters in France from Impressionism to Cubism. *The American Economic Review*, 91(4), 1063–1071.
- Gilbert, N. (1997). A simulation of the structure of academic science. *Sociological Research Online*, 2(2), [www.socresonline.org.uk/2/2/3.html](http://www.socresonline.org.uk/2/2/3.html).
- Goffman, W. (1966). Mathematical approach to the spread of scientific ideas. *Nature*, 212, 449-452.
- Goffman, W. & Harmon, G. (1971). Mathematical approach to the prediction of scientific discovery, *Nature*, 229, 103-104.
- Gould, R. (1991). Multiple networks and mobilization in the Paris commune, 1871. *American Sociological Review*, 56, 193-201.
- Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L.A.N. (2005). Team assembly mechanisms determine collaboration structure and team performance. *Science*, 308(29), 697-702.
- Guy, J., Gan, J., Selfridge, J., Cobb, S., & Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. *Science*, 315, 1143-1147.
- Hargadon, A. & Sutton R.I. (1997). Technology brokering in a product development firm. *Administrative Science Quarterly*, 42, 716-749.
- Hennessey, B.A., & Amabile, T.M. (2010). Creativity. *Annual Review of Psychology*, 61, 569-598.
- Kuhn, T.S. (1962). *The structure of scientific revolutions*. Chicago: Chicago University Press.
- Kurzberg, T.R., & Amabile, T.M. (2000). From Guilford to creative synergy: opening the black box of team-level creativity. *Creativity Research Journal*, 13, 285-294.
- Lambiotte, R., & Panzarasa, P. (2009). Communities, knowledge creation and information diffusion. *Journal of Informetrics*, 3, 180-190.
- Lane, D., Malerba, F., Maxfield, R., & Orsenigo, L. (1996). Choice and Action. *Journal of Evolutionary Economics*, 6(1), 43-76.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*, Milton Keynes: Open University Press.
- Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Thousand Oaks: Sage.
- Lotka, A. (1926). The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences*, 16, 317.
- March, J.G. (2007). The study of organizations and organizing since 1945. *Organization Studies*, 28, 9-19.
- Merton, R.K. (1973). *The sociology of science*. Chicago: University of Chicago Press.
- Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Newman M.E.J. (2004) “Coauthorship Networks and Patterns of Scientific Collaboration”, *Proceedings of the National Academy of Sciences*, **101**(1), 5200-5205.

- Orsenigo, L., Pamolli, F., & Riccaboni, M. (2001). Technological change and network dynamics: Lessons from the pharmaceutical industry. *Research Policy*, 30, 485-508.
- Owen-Smith, J., Riccaboni, M., Pammolli, F., & Powell, W.W. (2002). A comparison of US and European relations in the life sciences. *Management Science*, 48(1), 24-43.
- Padgett, J.F. & Ansell, C.K. (1993). Robust action and the rise of the Medici, 1400-1434. *American Journal of Sociology*, 98, 1259-1319.
- Padgett, J.F. & Powell, W.W. (2011). *The emergence of organizations and markets*, Princeton: Princeton University Press, forthcoming.
- Perry-Smith, J.E., & Shalley, C.E. (2003). The social side of creativity: A static and dynamic social network perspective. *Academy of Management Review*, 28(1), 89-106.
- Perry-Smith, J.E. (2006). Social yet reactive. the role of social relationship in facilitating individual creativity. *Academy of Management Journal*, 49(1), 85-101.
- Poincaré, H. (1921). *The foundations of science*. New York: The Science Press.
- Powell, W.W., Koput, K.W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41(1), 116-145.
- Rinia, E.J., van Leeuwen, Th.N., van Vuren, H.G., van Raan, A.F.J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria. *Research Policy*, 27, 95-107.
- Roth, C., & Cointet, J-P. (2010). Social and semantic coevolution in knowledge networks, *Social Networks*, 32, 16-29.
- Schumpeter, J.A. (1934). *The theory of economic development*. London: Oxford University Press.
- Shalley, C.E., & Perry-Smith, J.E. (2008). The emergence of team creative cognition: The role of diverse outside ties, sociocognitive network centrality, and team evolution. *Strategic Entrepreneurship Journal*, 2, 23-41.
- Simon, H.A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3-4), 425-440.
- Simon, H.A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6), 467-482.
- Simonton, D.K. (1988). *Scientific genius: A psychology of science*. Cambridge: Cambridge University Press.
- Simonton, D.K. (2000). Creativity: Cognitive, personal, developmental, and social aspects. *American Psychologist*, 55(1), 151-158.
- Sterman, J.D., & Wittenberg, J. (1999). Path dependence, competition and succession in the dynamics of scientific revolution. *Organization Science*, 10(3), 322-341.
- Taramasco, C., Cointet, J.-P., Roth, C. (2010). Academic team formation as evolving hypergraphs. *Scientometrics*, 85(3), 721-740.
- Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. *American Journal of Sociology*, 111, 447-504.
- Valente, T.W. (1995). *Network models of the diffusion of innovation*. Cresskill NJ: Hampton Press.

- Yin, R. (1994). *Case study research: design and methods*. Thousand Oaks: Sage.
- Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of 'Small-World' networks. *Nature*, 393, 440-442.
- Watts, D.J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the USA*, 99, 5766-5771.
- Weitzman, M.L. (1998). Recombinant growth. *Quarterly Journal of Economics*, 113(2), 331-360.
- White, H.D., Lin, X., Buzydlowski, J.W., Chen, C. (2004). User-controlled mapping of significant literatures, *Proceedings of the National Academy of Sciences of the USA*, 101, 5297-5302.
- White, H.C. (1993). *Careers and creativity: Social forces in the arts*. Boulder: Westview Press.
- Wuchty, S., Jones, B.F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316, 1036-1039.

## Appendix 1: Questions put to scholars on Rett Syndrome

1. Please write the names of the five most influential scientists (except your own) on Rett Syndrome, in order of the importance of their scientific contributions;
2. Please write the names of the five most creative scientists (except your own) on Rett Syndrome;
3. Please rank the five most important scientific publications (except yours) on Rett Syndrome (any citation style, but include at least: first author, year, title, journal/book editor);
4. Please list at least five key concepts which define the most important topics of research on Rett Syndrome.

## Figures and Tables

Table 1. Creativity in socio-semantic networks

Steps in creative problem-solving:	Problem identification	Conceptual combination	Conceptual expansion
Networks events:	New nodes	New links	Replication of links
Semantic network of concepts	Generation of new concepts <i>(generative creativity)</i>	Generation of new combinations of concepts <i>(combinatorial creativity)</i>	Replication of concept combinations, validation and diffusion <i>(replicative)</i>
Social network of scientists	Entry of new scientists (novices)	New scientific collaborations	Replication of collaborations

Figure 1. The distribution of new combinations of concepts/authors in the scientific papers on Rett Syndrome (source: PubMed, 2009)

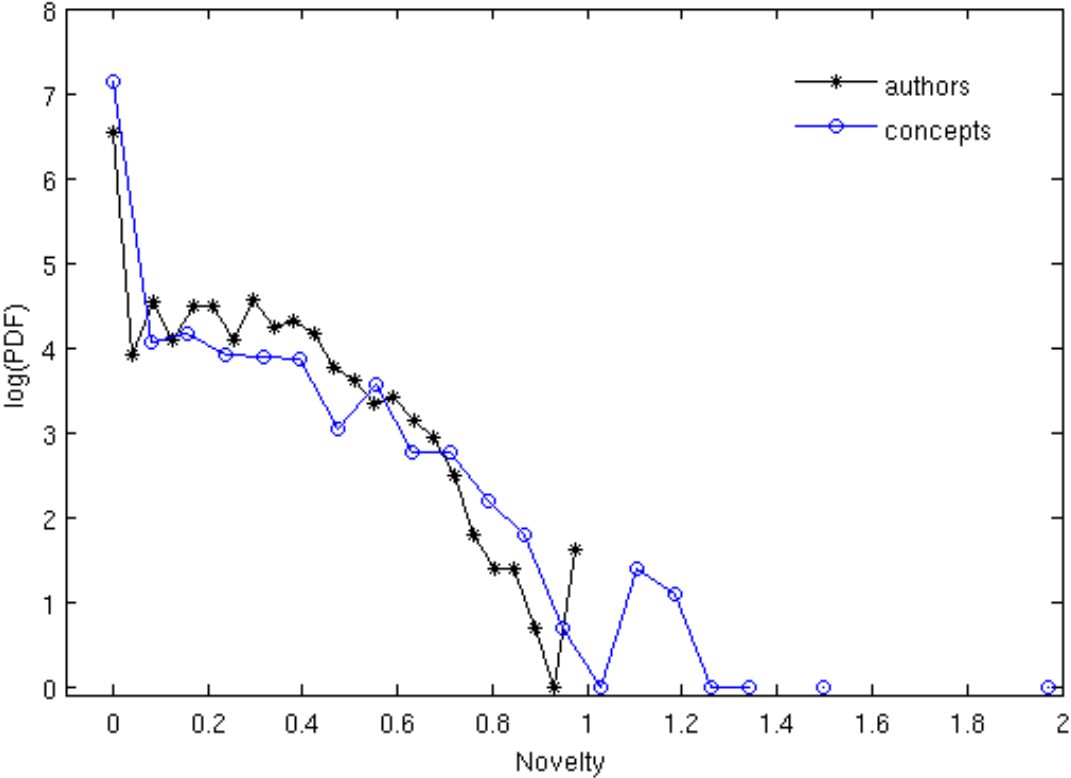


Figure 2. Average number of new concepts,  $n(c)$ , and the logarithm of the number of new combinations of concepts  $n(c^2)$  by authors' scientific experience as measured by number of articles previously appearing in PubMed

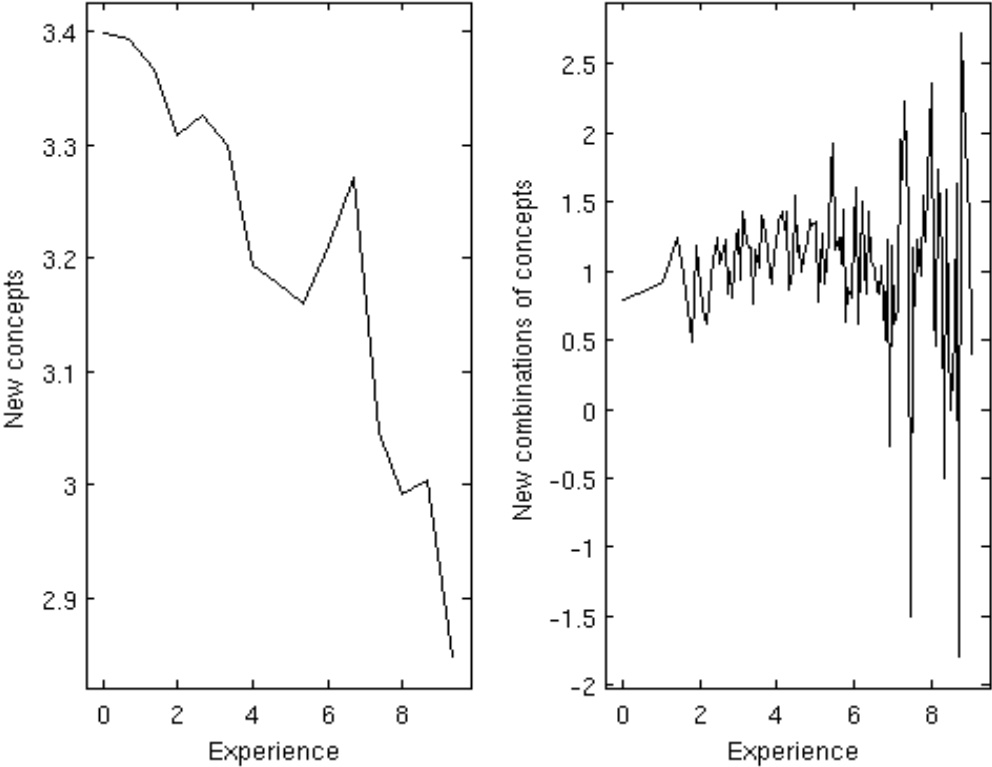




Figure 3. The logarithm of the number of times newly introduced (combination of) concepts have been replicated versus the logarithm of the centrality of the scientists in the scientific community of Rett Syndrome (source: PubMed, 2009)

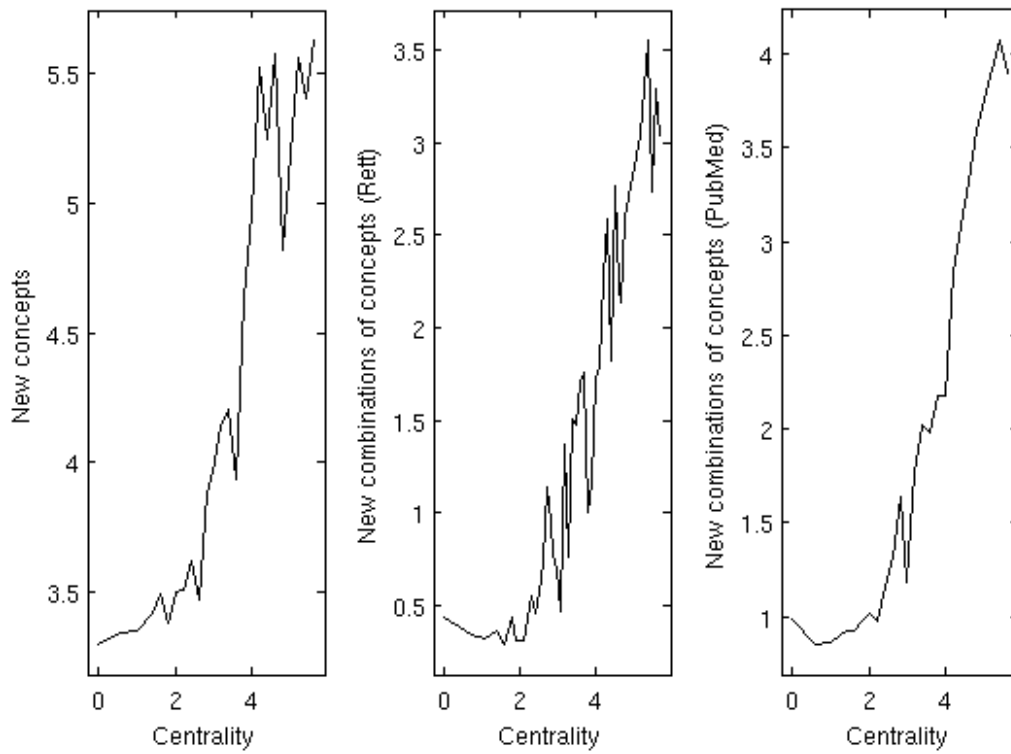


Table 2. Creativity in Rett syndrome scientific community: questionnaire results, bibliometric indicators and network statistics

KEY SCIENTIST <sup>1</sup>		QUESTIONNAIRE RESULTS									BIBLIOMETRIC INDICATORS					NETWORK STATISTICS				
		Creativity			Relevance			Key papers			TOTAL					RETT				
		Production <sup>2</sup>	h index <sup>2</sup>	Papers <sup>3</sup>	IF, total <sup>4</sup>	IF, yearly <sup>4</sup>	Strength	Degree	k-core	Entry										
Rett	Andreas	11	2.2	4	15.6	6	13.3	5	10.4	42	n.a.	4(4)	4.9	0.3	9	9	5	1952		
Hagberg	Bengt	10	3.3	5	15.0	8	6.6	8	8.3	110	12	46(21)	54.3	2.7	109	60	15	1983		
Percy <sup>5</sup>	Alan	10	3.3	9	3.8	--	--	15	3.6	151	16	43(25)	101.7	5.1	189	126	16	1989		
Zoghbi	Huda	2	24.6	2	61.3	2	54.7	2	46.9	261	64	51(36)	279.9	14.7	247	173	23	1990		
Naidu	SakkuBai	--	--	16	1.0	--	--	24	1.0	123	26	45(18)	150.2	7.9	242	172	20	1990		
Kerr	Alison	--	--	11	2.2	--	--	16	2.2	57	17	45(8)	79.6	4.2	271	190	25	1990		
Armstrong	Dawna	--	--	13	1.7	--	--	19	1.7	37	22	22(10)	49.5	2.6	82	73	20	1990		
Francke <sup>5</sup>	Uta	11	2.2	--	0.0	5	14.3	11	5.5	432	36	21(17)	105.7	5.9	93	68	20	1991		
Bird	Adrian	1	66.9	1	70.6	1	77.1	1	71.5	150	50	33(25)	313.7	18.5	113	82	17	1992		
Leonard	Helen	9	6.7	12	2.1	--	--	13	4.4	71	16	61(23)	93.4	5.8	425	205	20	1993		
Schanen	Carolyn	11	2.2	--	0.0	12	4.0	17	2.1	34	13	14(6)	65.9	5.5	73	67	13	1997		
Wade	Paul	--	--	--	0.0	--	--	--	--	64	33	15(7)	105.0	9.5	40	28	10	1998		
Landsberger <sup>5</sup>	Nicoletta	10	3.3	--	0.0	--	--	19	1.7	28	11	6(2)	52.0	4.7	49	39	19	1998		
Christodoulou <sup>5</sup>	John	13	1.3	--	0.0	--	--	25	0.7	152	27	47(17)	103.7	9.4	335	176	20	1998		
Renieri	Alessandra	--	--	13	1.7	--	--	19	1.7	122	25	23(13)	75.2	8.4	255	108	23	2000		
Bienvenu <sup>5</sup>	Thierry	--	--	--	0.0	--	--	--	--	148	25	19(8)	80.4	8.9	166	89	17	2000		
LaSalle	Janine	6	11.3	7	10.1	15	0.4	10	7.3	45	16	22(16)	86.4	10.8	75	46	11	2001		
Guy	Jacky	11	2.2	--	0.0	4	21.6	9	7.9	8	7	4(0)	43.2	5.4	23	18	10	2001		
Jaenisch	Rudolph	7	9.7	6	13.2	3	26.9	3	16.6	187	78	12(6)	104.5	13.1	61	47	10	2001		
Sun	Yi	4	18.3	9	3.3	8	8.2	6	10.0	32	21	7(4)	30.2	4.3	31	25	18	2002		
Mandel	Gail	5	12.8	8	9.4	10	4.7	7	9.0	28	18	3(1)	15.1	2.2	18	17	10	2002		
Eubanks	James	8	8.0	11	2.2	13	2.2	14	4.1	59	14	9(7)	21.7	3.1	35	20	6	2002		
Woodcock	C.L.	10	3.3	--	0.0	14	1.1	20	1.5	86	21	6(4)	35.7	5.9	24	12	5	2003		
Greenberg	Michael	3	21.3	3	16.6	7	9.2	4	15.7	120	64	5(4)	36.7	6.1	28	22	13	2003		
Chang	Qiang	13	1.3	--	0.0	11	4.4	18	1.9	13	9	4(0)	24.4	4.1	26	23	10	2003		
Moretti	Paolo	11	2.2	--	0.0	--	--	23	1.1	6	4	4(4)	8.6	2.1	24	20	10	2005		
Monteggia	Lisa M.	--	--	14	1.3	--	--	22	1.3	53	22	5(4)	17.9	6.0	13	9	5	2006		
Katz	David	9	6.7	10	2.7	--	--	12	4.7	94	18	6(4)	2.2	0.7	24	16	6	2006		
Pozzo-Miller	Lucas	12	1.7	15	1.1	--	--	21	1.4	45	19	3(3)	7.4	3.7	15	10	8	2007		
Justice	Monica	14	1.1	15	1.1	--	--	23	1.1	115	26	--	--	--	--	--	--	--		

<sup>1</sup> Scientists ordered by year of entry into RETT community; <sup>2</sup> Total number of papers, Source: Scopus; <sup>3</sup> PubMed papers on Rett (of which as the last author);<sup>4</sup> Source: our computation on ISI Web of Science; <sup>5</sup> Questionnaire Responders

Table 3. Creativity and entry in six rare diseases at different stages of evolution (Early: Horner and Noonan; middle: Rett and ADA; Advanced: PNH and Mesothelioma). Structural indicators (Number of nodes: Size; share of nodes with more than one link:  $F_R$ ; share of repeated links:  $f_R$ ; share of nodes in the giant component of the network:  $S$ ); maximum likelihood estimates of model parameters ( $p_1^a, p_1^f, p_2^a, p_2^f$ ) based on simulation results (sim.)

Rare disease	N. papers	Average number of:		Size	F_R	f_R	S	$p_1$	$p_1^a - p_1^f$	$p_2$	$p_2^a - p_2^f$		
1. Horner N. of trials <sup>1</sup> : 0 Prevalence <sup>2</sup> : n.a.	1,473	Authors	2.85	real	3,860	5.88	1.36	2.23					
				sim.	3,872	7.77	1.14	1.22	.85		0		
		Concepts	9.57	real	1,868	51.45	23.61	100					
				sim.	1,841	56.36	22.98	72.00	.07	.78	0	--	
2. Noonan N. of trials <sup>1</sup> : 3(0) Prevalence <sup>2</sup> : 50/100.000	985	Authors	4.70	real	3,466	14.08	7.63	32.60					
				sim.	3,455	20.49	5.58	22.31	.75		0		
		Concepts	10.40	real	1,591	43.87	23.91	100					
				sim.	1,566	45.58	21.95	100	.15	.60	0	--	
3. Mesothelioma N. of trials <sup>1</sup> : 143(45) Prevalence <sup>2</sup> : 2.7/100.000	9,625	Authors	4.04	real	23,748	22.21	18.54	31.54					
				sim.	23,756	27.21	19.21	29.81	.62		.05		
		Concepts	11	real	4,677	60.45	31.65	100					
				sim.	4,731	50.30	29.46	100	.04	.57	.05	--	
4. PNH N. of trials <sup>1</sup> : 25(9) Prevalence <sup>2</sup> : .55/100.000	2,476	Authors	3.78	real	5,849	19.94	11.74	29.71					
				sim.	5,833	29.00	11.73	31.00	.72		.02		
		Concepts	9.30	real	2,344	54.10	26.67	99.87					
				sim.	2,337	52.55	25.27	100	.10	.53	.02	--	
5. Rett N. of trials <sup>1</sup> : 13(4) Prevalence <sup>2</sup> : 4.15/10000	1,653	Authors	4.73	real	4,498	24.00	12.15	49.13					
				sim.	4,496	29.81	12.40	72.06	.57		.07		
		Concepts	12.23	real	2,283	52.00	26.81	100					
				sim.	2,274	48.25	24.19	100	.11	.46	.07	--	
6. ADA N. of trials <sup>1</sup> : 8(3) Prevalence <sup>2</sup> : .22/100.000	827	Authors	3.79	real	1,618	28.06	14.07	59.58					
				sim.	1,620	32.53	14.17	77.21	.52		.15		
		MESH terms	11.63	real	1,352	51.85	25.09	100					
				sim.	1,350	47.93	23.54	100	.14	.38	.15	--	

<sup>1</sup> Source: [www.clinicaltrials.gov](http://www.clinicaltrials.gov) (active trials into brackets). <sup>2</sup> Source: [www.orpha.net](http://www.orpha.net), report series n.2, November 2009





