Using Mixed Node Publication Network Graphs for Analyzing Success in Interdisciplinary Teams

André Calero Valdez[1], Anne Kathrin Schaar[1], Martina Ziefle[1], Andreas Holzinger[2], Sabina Jeschke[3], Christian Brecher[4]

[1] Human-Computer Interaction Center, RWTH Aachen University, Aachen, Germany
{calero-valdez,schaar,ziefle}@comm.rwth-aachen.de
[2] Institute for Medical Informatics, Statistics and Documentation, Medical University Graz
andreas.holzinger@medunigraz.at
[3] Institute of Information Management in Mechanical Engineering (IMA), Center for Learning and Knowledge Management (ZLW), Assoc. Institute for Management Cybernetics e.V. (IfU), RWTH Aachen University, Aachen, Germany
sabina.jeschke@ima-zlw-ifu.rwth-aachen.de
[4] Laboratory for Machine Tools and Production Engineering, RWTH Aachen University, Aachen, Germany,
c.brecher@wzl.rwth-aachen.de

Abstract. Large-scale research problems (e.g. health and aging, eonomics and production in high-wage countries) are typically complex, needing competencies and research input of different disciplines [1]. Hence, cooperative working in mixed teams is a common research procedure to meet multi-faceted research problems. Though, interdisciplinarity is – socially and scientifically – a challenge, not only in steering cooperation quality, but also in evaluating the interdisciplinary performance. In this paper we demonstrate how using mixed-node publication network graphs can be used in order to get insights into social structures of research groups. Explicating the published element of cooperation in a network graph reveals more than simple co-authorship graphs. The validity of the approach was tested on the 3-year publication outcome of an interdisciplinary research group. The approach was highly useful not only in demonstrating network properties like propinquity and homophily, but also in proposing a performance metrics of interdisciplinarity. Furthermore we suggest applying the approach to a large research cluster as a method of self-management and enriching the graph with sociometric data to improve intelligibility of the graph.

Keywords: Publication Network Analysis, Sociometry, Interdisciplinarity, Research Cluster Assessment, Bibliometry, Visualization

1 Introduction

Interdisciplinarity is a hyped term when it comes to directions of scientific research [2]. Inter- or transdisciplinary approaches promise breakthrough developments [3] by leveraging method competences from different fields in unison. Scientific teams have been shown to outperform solo authors in knowledge generation[4].

In order to acquire funding for research scientists often need to look into interdisciplinary approaches to solve real world problems. But interdisciplinarity cannot be achieved by simply combining researchers from different fields into a research group. In contrast, interdisciplinarity – though widely acknowledged as a reasonable research procedure from a technical point of view – suffers from diverse cognitive research models across team members, stemming from different knowledge domains, research languages, methods, models, and procedures. Aggravating, as team members are mostly not aware of threes different professional upbringings, team's cooperation is often not perceived as successful or effective by team members [5].

Efforts have been made to understand how interdisciplinarity must be learnt from a socio-cultural, social, cognitive perspective to gain insights on the learning processes of interdisciplinarity as a faculty [6]. It has been found that successful interdisciplinarity requires a conscious effort, time and resources to establish the required interpersonal relationships for effective communication [7]. Successful teams have also been shown to perform better at interdisciplinarity than newly formed teams [8].

But before one can select measures to improve communication effectiveness or interpersonal relationships it is necessary to determine what factors contribute to interdisciplinary success and furthermore what constitutes interdisciplinary success.

In traditional disciplinary research established and widely accepted methods of measuring success exist. But how can one translate measurements like the judgment of an established community for peer review if no established group of peers exists. Quality of outlets by measuring impact factors might also be inappropriate, because young interdisciplinary fields of research have no established outlets, and acclaimed disciplinary focused outlets might reject interdisciplinary publications due to misunderstanding or out of scope problems [9].

Assuming that publications are a measure of disciplinary success, publication cultures differ between disciplines leaving interdisciplinary research without a unified calibrated measure for success.

1.1 Using Publication Network Analysis to Manage Success

Understanding how families of scientific disciplines differ has already been analyzed by Publication Network Analyses [10]. Also flows of citations have been used to analyze development of a research field [11] in highly inter- and transdisciplinary field. Web-Based Data mining of publication data can be used to understand how scientific fields progress [12]. Using graph representation for publication analysis suggests itself because of the innate graph-like structure of publications. Inbetweenness Centrality of Journal Graphs has been used as a measurement for interdisciplinarity in outlets [13].

Even if publications are a valuable measuring tool for whole fields of research, how can one identify latent structures that lead to high quality scientific output in specific interdisciplinary teams. Understanding how groups of people are linked and how they can be affected has been studied in the early 50ies with sociometry [14].

Mapping qualitative data (e.g. who talks with whom) to graphs reveals important nodes and possible change agents to influence the whole social network.

But what are the implications for interdisciplinary teams? Can one do measuring and steering interdisciplinary research efforts by looking as sociometric data and publication networks [15]? Do similarities exist?

The idea of the quantified self [16] defines a new perspective that uses specific (mobile) applications for measuring parameters (vital or habitual) in order to allow self-management. Whenever something is measured intentionally, the outcome is altered during the measurement (by the awareness for the measurement). This effect is often applied in cognitive behavioral therapy by increasing awareness of the measured dimension. This improved awareness increases self-efficacy and thus improvement in behavior[17]. Can this approach be used to allow steering of research groups?

2 Visualizing Publication Networks

The idea for using mixed node graphs for publication network analysis came to us when trying to demonstrate the research efforts of a highly active interdisciplinary research group at RWTH Aachen University (http://www.humtec.rwth-aachen.de/ehealth). The group and its research program started in 2009 (funded by the excellence initiative of German federal and state governments). In order to make research efforts and its success transparent to the German Wissenschaftsrat (the highest scientific board in Germany), we tried to understand how we have worked, why we were successful and what had lead to this development. For this purpose we generated a visualization of our publication behavior. But in order to see the interdisciplinary efforts, we needed something different than simple co-author networks, because the output of the cooperation (namely the publication) should be a part of the representation as well. The typical binary co-authorship relationship actually represents an n-ary relationship between n-1 authors and a publication. This is why we tried to use mixed node publication network graphs.

Graph theoretical analyses of bibliometric data usually use single node type network graphs (i.e. all nodes are authors or all nodes are publications). These mostly contain single typed edges (e.g. co-author relationship or citations). The use of mixed node publication network graphs allows a graph to contain more information (than a co-authorship graph) and can easily be reduced to one by using an injective mapping function. Making these entities part of the graph makes visual interpretation easier.

2.1 How the Mixed Node Publication Network Graph Is Constructed

The network graph G is constructed with mixed node types. A node either represents an author (A-Node) a publication (P-Node) or a discipline (D-Node). From a graph theory point of view nodes (i.e. vertices) are not regarded as differently. We define three sets representing authors, publications and disciplines:

$$A = \{\, a \mid a \text{ is author in ehealth research group}\} \qquad (1)$$

$$P = \{p \mid p \text{ is a publication written by any } a \in A\} \tag{2}$$

$$D = \{d \mid d \text{ is a discipline studied by any } a \in A\} \tag{3}$$

Then we can define three vertex-mappings f_a, f_p and f_d and three sets of vertices V_1, V_2 and V_3 as follows:

$$f_a : A \rightarrow V_1,\ f_a(a) = v;\ a \in A \wedge v \in V_1 \tag{4}$$

$$f_p : P \rightarrow V_2,\ f_p(p) = v;\ p \in P \wedge v \in V_2 \tag{5}$$

$$f_d : D \rightarrow V_3,\ f_d(d) = v;\ d \in D \wedge v \in V_3 \tag{6}$$

$$\text{with } V_1 \cap V_2 \cap V_3 = \varnothing \tag{7}$$

We define the sets E_1 and E_2 and a weight mapping ω as follows, using that f^{-1} is the inverse of f:

$$E_1 = \{\ e \mid e=(v_1, v_2),\ v_1 \in V_1 \wedge v_2 \in V_2 \wedge if\ f_a^{-1}(v_1) \text{ is author of } f_p^{-1}(v_2)\} \tag{8}$$

$$E_2 = \{\ e \mid e =(v_1, v_3),\ v_1 \in V_1 \wedge v_3 \in V_3 \wedge if\ f_a^{-1}(v_1) \text{ studied discipline } f_d^{-1}(v_3)\} \tag{9}$$

$$\omega : E \rightarrow \mathbb{R},\ \omega(e) = 1, \text{ if } e \in E_1 \text{ and } \omega(e) = 0.5, \text{ if } e \in E_2 \tag{10}$$

Then we define two graphs as follow: G_r we call the reduced mixed node publication network graph and G_f we call a full mixed node publication network graph.

$$G_r = (V, E) \text{ with } V = V_1 \cup V_2 \text{ and } E = E_1 \tag{11}$$

$$G_f = (V, E), \text{ with } V = V_1 \cup V_2 \cup V_3 \text{ and } E = E_1 \cup E_2 \tag{12}$$

The reduced and full mixed node publication network graphs are representations of publication networks that can be visualized using standard graph visualization tools. G_r is a bipartite and G_f a tripartite graph.

2.2 Spatial Mapping of the Publication Graph

In order to allow visual analysis by a human person graphs need be lain out graphically. For this purpose we use the open-source software Gephi [18]. Gephi allows graph input by various means (e.g. HTTP-JSON interface) and different layout algorithms.

In this case 2D-spatial mapping is performed by Gephi using it's *Force-Atlas 2* algorithm. Graphs in Gephi allow additional information for graph elements. In particular color, size and labeling can be defined for edges and vertices (i.e. nodes).

For our visualization we set the size of P-Nodes to 10, A-nodes to 50 and D-Nodes to 100 (see fig. 1).

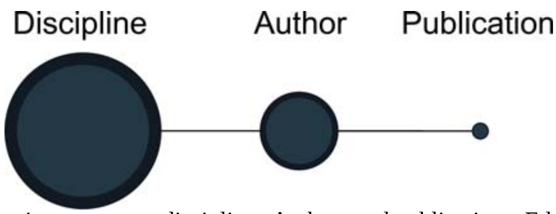

Fig. 1. Three node sizes represent disciplines Authors and publications. Edges represent relationships between nodes.

Using the *Force-Atlas 2* algorithm creates a visual representation according to the following rules:

- All nodes are attracted to the center.
- All nodes repel each other.
- All nodes that are connected by an edge attract each other, according to the weight of the edge (i.e. heavier edges equals stronger attraction).
- Optionally node sizes can be added to the repulsion to prevent visual overlapping.

This allows the following visual conclusions:

- Two A-Nodes are spatially closer if they publish together.
- Two A-nodes are spatially closer if they come from the same discipline.

2.3 Temporal Mapping of Graph

In order to understand development of publication networks temporal mappings or snapshots were required. The idea was that according to publication date nodes and edges were sequentially entered into the graph using the JSON interface of Gephi. This would allow programmatically based animated publication graphs.

Timing of insertion is structured (pauses between new years) to give the impression of stretching time, which allows the layout algorithm to further sort nodes spatially. This sorting and inserting is recorded into a video file. The resulting video of the sorting algorithm is then sped up until it fits into a 90-second clip.

In this clip nodes move according to the attractive forces of the continuously running layout-algorithm giving the impression of a birds-eye-view of moving people that group together. The human brain (even in its early infancy) tends to apply agenticity[19] (infer agents behind patterns) if objects move in atypical non-physical motions [20]. This further enhances the impression of persons moving to find their “peer group” in the publication network graph (see https://vimeo.com/48446978).

2.4 Benefits of Visually Mapping Mixed Node Publication Networks

Gephi allows for several different graph analyses of network graphs. Traditionally these are used with social network graphs (i.e. co-authorship graphs). Interpretation of graph statistics must be reevaluated for mixed node graphs. Graph statistics that are of interest in regard to publication networks are:

- *Number of Weakly Connected Components* [21] refers to the amount of components that are only weakly connected (i.e. only by directed edges in one direction). In an undirected graph they reflect the number of unconnected communities (i.e. subgraphs).
- *Graph density* reflects to the degree of how connected a network graph is. If the density is 1 all nodes are connected with each other. Higher density means that the network is better connected. For bipartite graphs (like G_r) maximal density is limited by:

  $11+ \quad -12 \quad + \quad -12$, when *n*,*m* are the cardinalities of the two parts.

- *Graph Diameter* refers to the maximal distance between any two nodes in a network. The smallest possible diameter is 2 (for G_r) and 3 (for G_f). When more than one discipline exists in a graph the smallest possible diameter can become 5 (if two authors of two different discipline publish tighter). Larger diameters mean that some authors in the network are not publishing together.
- *Average Path Length* refers to the average length from any node to all other nodes. Larger numbers can mean less cooperation or the existence of highly central nodes (that lie on many paths). It cannot be lower than 1 (for G_r) and 2 (for G_f).
- *Average Degree* refers to the average of outgoing edges in the graph, represents the average of publications per author mixed with the average of authors per publication. When using G_f one must be aware of the two confounding influences. The average number of authors per discipline and the average number of disciplines per author. This makes immediate interpretation of this value harder.
- *Betweenness Centrality, Closeness Centrality, Eigenvector Centrality and Eccentricity* [22] are measures for nodes indicating how important they are for finding short paths in the network. The Closeness Centrality reflects the average importance of a node when randomly spreading information to the whole network (which might be used to model communication flow), while Betweenness Centrality reflects the average importance of a node to find a shortest path between two specific nodes. Eigenvector Centrality measures the importance of a node for the total network. Central persons (i.e. Professors) should show high values in Betweenness Centrality, Closeness Centrality and Eigenvector Centrality. Eccentricity refers to the maximum possible distance to any other node for a specific node. It can only be smaller than the diameter and should be high in weakly connected nodes.
- *Modularity and Community Detection* [19-20] can be used to identify groups in connected graphs that share more edges than randomness would predict. Modularity then measures the amount of how much higher the connections within a community are against connections between communities. Lower values mean that communities interact more with another.

The human mind is capable of analyses that are not computationally easy. Tasks that are relatively easy for the human brain but hard for computers are called “Human

Intelligence Tasks" (HIT). Seeing structural changes in a network graph from a meta-perspective is one of those tasks (e.g. seeing whether two subgraphs are connected).

Especially interpreting measures like density and centrality is rather hard for mixed node graphs. Visualization makes the interpretation of these measures fundamentally easier. Enriching visualization with qualitative sociometrical data allows for high quality educated guessing in understanding a mixed node publication network graph.

3 Analyzing the Publication Network Visualization of the eHealth Research Group at RWTH Aachen University

Two types of analyses are possible: Graph Statistics from Gephi and informed pattern recognition from humans. Both are performed here as an example. As graph data publication data from the ehealth group is used as a full mixed node publication network graph. The term informed is used in this case because social anatomy of the group is well known by the author. The mixed node graph is shown here (see fig. 2). Furthermore nodes are colored according to the discipline to that they belong.

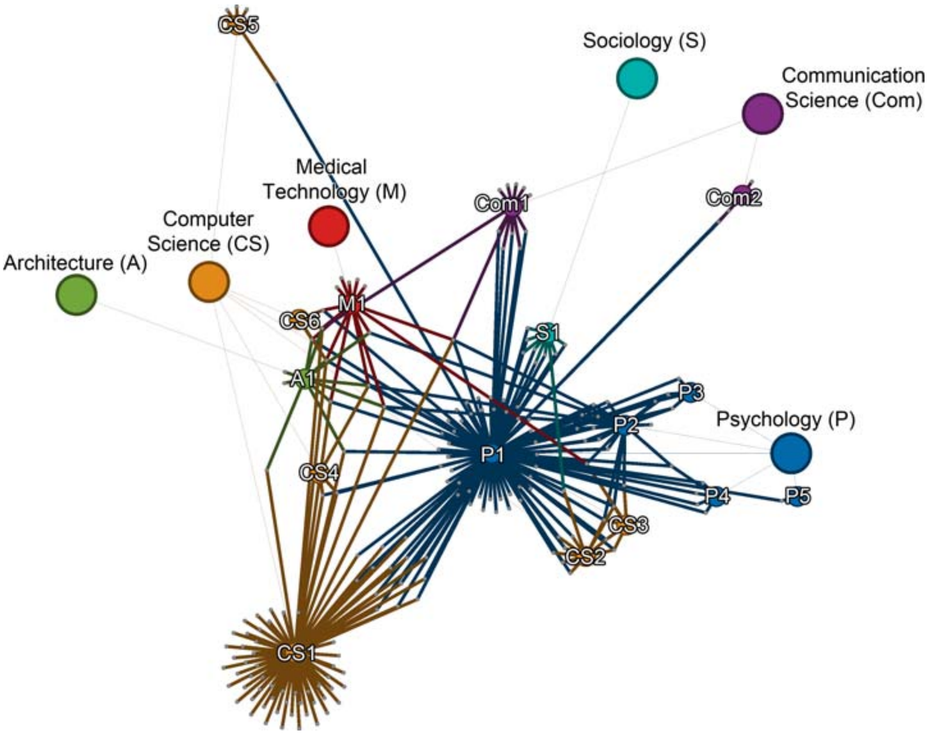

Fig. 2. G_f for the ehealth programme publication data over 3 years. Six disciplines, fourteen authors and 198 publications, (http://www.humtec.rwth-aachen.de/ehealth).

3.1 Graph Statistics

Applying the Gephi graph analysis reveals the following statistics. The graph contains 14 authors, 6 Disciplines and 198 publications. The average Degree is 3.009 and the diameter of the graph is 6. The average path length is 3.055 Graph density is .014. The graph only contains one weakly connected component, which has 8 communities and a modularity of .512. These results demonstrate a highly interconnected network with short paths between disciplines, authors and publications. In regard to centrality measures (closeness, inverse eccentricity, betweeness and eigenvector), two nodes are prominent P1 (1st place in all measures) and CS1 (2nd place in all measures). Nonetheless P1 and CS1 are dramatically different, as presented in the next section.

3.2 HIT-Analysis

When looking at the animated network graph certain additional factors become obvious, that are hard to see from the statistics point of view. Certain structures become visible which remain hidden from centrality measurements.

In this graph it is obvious that the node P1 plays a structurally important role, which is also predicted by the centrality measures. The node CS1 in contrast is predicted to play an important role, but visually remains on the outskirts of the graph. Looking at the social anatomy of the group reveals why CS1 is not located at the center. The person behind CS1 has had only bi-weekly attendance at the institute, and sits in a single-person office.

Typical social structures reveal themselves in a graph like propinquity and homophily assuming an underlying implicit multiplexity of the edges. Nodes that cluster together come from the same projects (e.g. Com1, A1, CS6, M1, P1), share offices (e.g. CS2, CS3, P2), come from similar discipline (e.g. P1, P2, P3, P4, P5), are friends in their free time (e.g. CS2, CS3) or apply similar methods in their research (e.g. CS2, CS3, P2). Interdisciplinary publication success becomes also visible by looking at the color distribution of the graph. Particularly the group of CS6, M1 and A1 have published very interdisciplinary.

Nodes that are also the outskirts of the graph (COM2 and CS5) are members of the team that have joined our team quite recently.

3.3 Additional Insights

One could have expected that information visualization would have evoked negative and competitive feelings within the group. However, the contrary was the case. When demonstrating the visualization within the group reactions where positive throughout. Not a single member of the team focused on ranking member into a publishing-top-list or anything similar. In contrast members of the group were astonished to see how their publication behavior was so revealing about themselves. Thus, the visualization did evoke additional interest for the group and a hedonic gaming attitude on how to increase interdisciplinary publication behavior as a mean for further team cooperation. For example, some members firstly realized that there are members of the team that shared research interests with them, but have not published together yet. Looking at publications from a revealing of existing and unpublished insights point of view, proved itself to be very helpful. Members reported the visualization be a motivating factor for themselves. This shows that (1) information visualization in form of picturing publication networks can facilitate social behavior and increase team identity and (2) performance measurement does not provoke hostile team behavior, if the reason for the performance visualization is made transparent and if the tool can be used as a self-control instrument of the group (rather than by heads only).

4 Cybernetic Application of Publication Network Visualization for Interdisciplinary Innovation Management

This lead to the question whether one can apply this approach in a cybernetic way to allow self-measurement to steer a scientific cluster? In order to test this idea, we first created a reduced mixed node publication network graph for the publications of the cluster of excellence (Integrative Production Technology for High-Wage Countries: http://www.production-research.de). The reduced graph was chosen, because no author information on disciplinarity was publically available (see fig. 3).

Applying the Gephi graph analysis reveals the following statistics. The graph shows an average degree of 3.766 and a Diameter of 23. The average path length is 8.08. Graph density is .005 (maximal theoretical possible density of this graph ~.18). Community analysis reveals 28 Communities and a modularity of .844.

In regard to node statistics two professors' nodes (located in the center of the graph) dominate the centrality measures with one exception. In regard to eigenvector centrality a node from the node-cluster in the top center ranks third. This node is a bridge node that has many strong ties within his group but also weak ties (which are important for allowing information between node-clusters) to another group.

One must wonder whether social analysis of this graph is possible? From various sources we have heard that the just reported bridge-node is also a person that is seen as interested in various topics, communicative and extroverted. This hint might lead to the conclusion that social structures are hidden in a graph but need to be studied on their won. This graph can only be analyzed and interpreted correctly if underlying social parameters are assessed. This could allow analyzing success factors of central nodes on the fly and allow steering by identifying networking agents or designing cluster specific seminars to enhance interconnectivity within a research cluster.

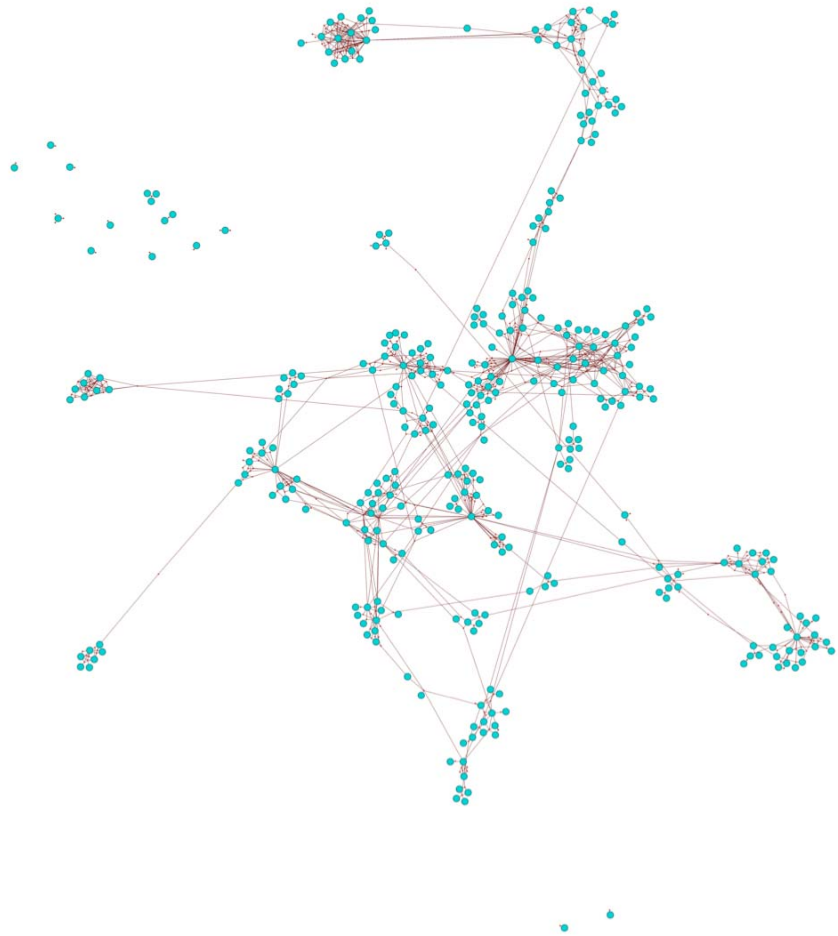

Fig. 3. G_r of publication data of a research cluster over five years of over 2500 publications and 274 Authors. Structures emerge immediately to the human eye.

5 Conclusion

Success factors for interdisciplinary research efforts can be measured by looking at publication network graphs. By using mixed node graphs important real world properties are added to the graph, which simplify human interpretation, by making implicit relationships (i.e. co-authorship) explicit (by showing co-authored work).

But in order to give publication network graphs more meaning further data is required. Parameters like impact factors, citation indices should be incorporated from a bibliometrical point of view. From a sociometric point of view properties like personality traits, motivation types, method competences and many more need to mapped to gain further insights.

From a graph theory point of view using mixed node graphs might break the interpretability of some of the used graph statistics but the enhancement of visibility outweighs this problem for the time being.

Acknowledgements: We would like to thank the members of the "ehealth research group" for allowing us to analyze their publication data. Furthermore we thank Dr. Ulrich Prahl and Eva Nick for generously supplying us with validated publication data of the CoE. This research was funded by the Excellence Initiative of the German federal and state governments.

References

1. Ziefle, M., Röcker, C., Wilkowska, W., Kasugai, K., Klack, L, Möllering, C. & Beul, S. (2011). A Multi-Disciplinary Approach to Ambient Assisted Living. In: Röcker, C., Ziefle, M. eds: E-Health, Assistive Technologies and Applications for Assisted Living: Challenges and Solutions, pp. 76–93 IGI Publishing, Niagara Falls, NY (2011).
2. Jacobs, J.A., Frickel, S.: Interdisciplinarity: A Critical Assessment. Annual Review of Sociology. pp. 43–65 Annual Reviews, Palo Alto (2009).
3. Bordons, M., Zulueta, M.A., Romero, F., Barrigón, S.: Measuring interdisciplinary collaboration within a university: The effects of the multidisciplinary research programme, Scientometrics, vol. 46, Nov. 1999, pp. 383–398.
4. Wuchty, S., Jones, B.F., Uzzi, B.: The Increasing Dominance of Teams in Production of Knowledge, Science, vol. 316, May. 2007, pp. 1036–1039.
5. Ziefle, M., Jakobs, E.M.: New challenges in Human Computer Interaction: Strategic Directions and Interdisciplinary Trends. Proceedings of the 4th International Conference on Competitive Manufacturing Technologies. pp. 389–398 University of Stellenbosch, South Africa (2010).
6. Lattuca, L.R.: Learning interdisciplinarity - Sociocultural perspectives on academic work, Journal of Higher Education, vol. 73, Dec. 2002, p. 711–+.
7. Marzano, M., Carss, D.N., Bell, S.: Working to make interdisciplinarity work: Investing in communication and interpersonal relationships, Journal of Agricultural Economics, vol. 57, Jul. 2006, pp. 185–197.
8. Cummings, J.N., Kiesler, S.: Who collaborates successfully?: prior experience reduces collaboration barriers in distributed interdisciplinary research. Proceedings of the 2008 ACM conference on Computer supported cooperative work. pp. 437–446 ACM, New York, NY, USA (2008).
9. Wilson, E.V., Lankton, N.K.: Interdisciplinary research and publication opportunities in information systems and healthcare, Communications of the Association for Information Systems (Volume 14, 2004), vol. 332, 2004, p. 343.
10. Grahl, J., Sand, B., Schneider, M., Schwind, M.: Publication Network Analysis of an Academic Family in Information Systems. In: Heinzl, A., Buxmann, P., Wendt, O., and Weitzel, T. (eds.) Theory-Guided Modeling and Empiricism in Information Systems Research. pp. 1–13 Physica-Verlag HD (2011).
11. Bassecoulard, E., Lelu, A., Zitt, M.: Mapping nanosciences by citation flows: A preliminary analysis, Scientometrics, vol. 70, Mar. 2007, pp. 859–880.

12. Cronin, B.: Bibliometrics and beyond: some thoughts on web-based citation analysis, Journal of Information Science, vol. 27, Feb. 2001, pp. 1–7.
13. Leydesdorff, L.: Betweenness centrality as an indicator of the interdisciplinarity of scientific journals, Journal of the American Society for Information Science and Technology, vol. 58, Jul. 2007, pp. 1303–1319.
14. Moreno, J.L.: Foundations of sociometry: An introduction, Sociometry, 1941, pp. 15–35.
15. Agasisti, T., Catalano, G., Landoni, P., Verganti, R.: Evaluating the performance of academic departments: an analysis of research-related output efficiency, Research Evaluation, vol. 21, Mar. 2012, pp. 2–14.
16. Wolf, G., Carmichael, A., Kelly, K.: The quantified self, TED http://www. ted. com/talks/gary_wolf_the_quantified_self. html, 2010.
17. Strecher, V.J., DeVellis, B.M., Becker, M.H., Rosenstock, I.M.: The Role of Self-Efficacy in Achieving Health Behavior Change, Health Education & Behavior, vol. 13, Mar. 1986, pp. 73–92.
18. Bastian, M., Heymann, S., Jacomy, M.: Gephi: An open source software for exploring and manipulating networks, 2009.
19. Shermer, M.: Agenticity, Scientific American, vol. 300, 2009, pp. 36–36.
20. Premack, D.: The infant's theory of self-propelled objects, Cognition, vol. 36, Jul. 1990, pp. 1–16.
21. Tarjan, R.: Depth-first search and linear graph algorithms. Switching and Automata Theory, 1971., 12th Annual Symposium on. pp. 114–121 (1971).
22. Brandes, U.: A faster algorithm for betweenness centrality*, Journal of Mathematical Sociology, vol. 25, 2001, pp. 163–177.
23. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, 2008, p. P10008.
24. Lambiotte, R., Delvenne, J.C., Barahona, M.: Laplacian dynamics and multiscale modular structure in networks, Arxiv preprint arXiv:0812.1770, 2008.