

The Risks of Introspection: A Quantitative Analysis of Influence Between Scientific Communities

Václav Belák and Conor Hayes

Insight Centre for Data Analytics

NUI Galway

Ireland

email:conor.hayes@nuigalway.ie

Abstract

Research impact is increasingly evaluated by quantitative analysis of citations, such as the h-index or conference impact factor. However, little attention has been given to understanding the patterns of influence and dependence that exist between groups of *scientific communities*. We propose that an analysis of inter-community influence offers valuable insights into how scientific communities evolve in terms of growth, stability, decline, information exchange, and impact. We present a computational model for inter-community influence in science and evaluate it on 19 years of data from communities in the field of Artificial Intelligence (AI). We uncover and explain the dynamics of communities that are *gaining* influence; that act as bridging *hubs*; or *isolated* communities that are apparently losing influence. This analysis reveals factors that may underpin a scientific community's growth or decline.

Introduction

There has been extensive research on the use of citation measures to evaluate the impact of individual scientists, institutions, or even entire countries (Moed 2005). These methods typically measure the *overall* impact an individual or an institution has on the surrounding research activities. However, there has not been much attention on understanding the patterns of influence and dependence that exist between groups of *scientific communities*.

Understanding the patterns of influence between *communities of scientists* provides a better contextualisation of the notion of scientific impact by observing the organisational structures in which scientific discourse is produced.

In this paper, we show how the growth, stability, impact and decline of a scientific community is predicated upon the relationships and *information exchange* (Goldstone and Leydesdorff 2006) it maintains with other communities.

We extend a model for inter-community influence. (Belák, Karnstedt, and Hayes 2011; Belák, Lam, and Hayes 2012), so that it captures the influence relations between scientific communities. We then evaluate the model on communities in the Artificial Intelligence (AI) field.

The main **contributions** are:

- A computational model for influence between scientific communities.
- An evaluation of the model on 19 years of data from conferences in the AI field of computer science.
- An analysis of the features that define community influence such as introspection, in-flow and out-flow

We review previous research in Section 2. The inter-community influence model COIN is presented in Section 3. We describe the data that we use to evaluate the model in Section 4. We proceed with the evaluation in two steps. In Section 5, we compare the behaviour of AI communities over 19 years. Then, in Section 6 we select one particular community for more in-depth qualitative and quantitative analysis to shed more light on the observed trends. We conclude and discuss the results in Section 7. Finally, all the software, data, and other outputs that are not presented in this paper for technical or other reasons are available as on-line supplementary material ¹.

Background

Citations between papers, books, authors, or institutions have been commonly used as a proxy for research impact (Moed 2005). Perhaps the most popular measures are 2-year journal (Moed 2005, p. 92) or conference (Martins et al. 2010) impact factor that measure the average number of citations a journal or proceedings of a conference receive in a particular year from papers published within the previous two years. Citations may thus be understood as a proxy for “*actual* influence on surrounding research activities at a given time” (Martin and Irvine 1983). Another way to estimate the impact of a citation is to directly measure how much information “flows” from the cited to the citing paper as depicted in Figure 1. This flow can be measured by how much language has been adopted by the citing paper (Dietz, Bickel, and Scheffer 2007). The ability of a research community to attract new researchers may be used as an indicator of the community's performance (Montolio, Dominguez-Sal, and Larriba-Pey 2013). In addition, new members provide access to ideas and resources from other communities. Likewise, a group of researchers is more likely to have impact if it is familiar with research outputs *beyond* its own

¹See <http://belak.net/doc/2015/flairs.html>.

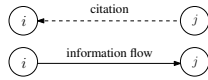


Figure 1: A citation from author j to author i indicates an explicit engagement of author j with a paper from author i . It suggests that author j ingested the content of the paper and thus that some information “flowed” in the opposite direction from author i to j .

boundaries (Jones, Wuchty, and Uzzi 2008). In this context (Goldstone and Leydesdorff 2006) talk about *import* and *export* of a community, i. e. the flow of ideas and knowledge to and from the community.

In short, the impact of a community is influenced by the position of the community in the network and by its relations with other communities. Our aim is to quantify and analyse the patterns of influence relations that communities in science may exhibit.

Community Influence Model

This section presents the main concepts of the model for inter-community influence, COIN, that has previously been used in the analysis of online Web communities (Belák, Lam, and Hayes 2012). However, the original model assumed that interactions occur in threads that belonged to single communities. If we consider a citation as a type of response to a previous paper, this constraint is not meaningful. We therefore ease this limit by proposing a more flexible representation based on tensors. First, we define the membership, centrality, and impact formally. After that, we derive several aggregate measures that we use in our analysis. We consider a general case of n authors participating in k communities (conferences).

Membership Matrix A set of authors attending a conference may be perceived as a community corresponding to some (sub-)discipline of science (Biryukov and Dong 2010). The distribution of an author’s publications over the conferences thus expresses the degree of her membership in each of the communities (Patil, Liu, and Gao 2013). We therefore define the $n \times k$ **membership matrix** \mathbf{M} representing a membership of author i in community u as: $M_{iu} = |P_{iu}| / \sum_x |P_{ix}|$, where P_{iu} is a set of papers contributed by author i to venue u within a time-window whose length is discussed later. A column $\mathbf{M}_{\cdot u}$ represents the memberships of the authors from community u , that is, the fuzzy set of the members of community u . The sum over the column thus represents the cardinality of the set or **size** of the community.

Centrality Tensor In order to apply COIN to communities of researchers, it is necessary to measure the centrality of each researcher within each community. We interpret a paper with a high number of citations as a paper that has had impact on the work of other scientists. If we consider a network of researchers connected by citations to their papers, a researcher’s raw in-degree is an indication of their overall influence.

We therefore consider a researcher’s impact within any community as a function of their in-degree in that community. As a researcher may be cited from multiple communities, we measure the author’s centrality with respect to all communities.

The $n \times k \times k$ **centrality tensor** $\mathcal{C} : \mathcal{C}_{iuv}$ representing a centrality of author i in community v due to her publications in community u is defined as the total number of citations from papers published at v to the papers published by author i at conference u . Therefore, the centrality may be interpreted as a tendency of author i to stimulate responses (citations) from the members of community v within the time-window. In the case a paper has multiple authors, we assign its citations to each of the co-authors (i. e. we adopt integer counting (Moed 2005, p. 273)), because in the data available to us there is no quantitative accounting of credit of the individual co-authors.

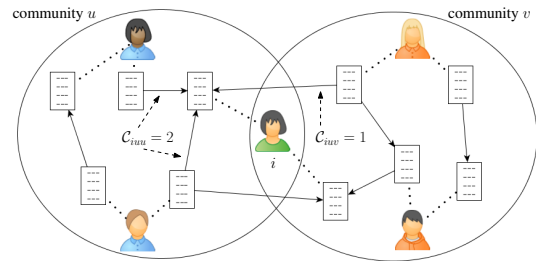


Figure 2: An illustration of the citations-based centrality of author i (green) at the intersection of the two communities. The undirected dotted links represent authorship, the directed solid links represent citations, and the dashed directed links depict the citations that contribute to the author’s centrality.

Figure 2 illustrates our definition of citations-based centrality. We see that author i contributed one paper to each of the communities u and v . Because her paper from community u was cited two times by the other papers from the same community, the centrality of author i in community u due to her publication in the same community is $\mathcal{C}_{iuu} = 2$. Analogously, since the same publication by author i from community u received one citation from a paper from community v , the centrality of author i in the other community v due to her publication in u is $\mathcal{C}_{iuv} = 1$.

Impact Matrix Analogously to the original model (Belák, Lam, and Hayes 2012), we define the cross-community impact \mathbf{J}_{uv} of community u on v as a *mean* centrality of the members of community u in community v , *weighted* by their membership in u :

$$\mathbf{J}_{uv} = \frac{\sum_x^n \mathbf{M}_{xu} \mathcal{C}_{xuv}}{\sum_{x=1}^n \mathbf{M}_{xu}}, \quad (1)$$

where the divisor corresponds to the size of community u . All impacts between any pair of communities can be conveniently represented as a $k \times k$ **impact matrix** $\mathbf{J} : \mathbf{J}_{uv}$, whose elements may be interpreted as a tendency of the members of v to cite the members of u . As illustrated by Figure 1m

citations may be interpreted as an indicator of (reverse) information flow. Therefore the impact of u on v may be interpreted as a measure of *information flow* from u to v .

The diagonal elements \mathbf{J}_{uu} represent the self-impact of the community, i.e. the degree to which a community refers to its own results. We refer to this impact \mathbf{J}_{uu} as the **introspection** of community u .

Aggregate Measures As well as measuring the influence between pairs of communities, we also define a set of *aggregate measures* that quantify the *overall* influence between a community and the other communities. Since we interpret the impact as a proxy for information flow between two communities, the aggregated measures indicate the overall flow from and to the community.

While the distribution of the impact values over one row of \mathbf{J} is useful for low-level cross-community analysis, the sum over the row represents the overall *out-flow* from the community to the other communities. In order to account only for the impact a community has on others we exclude the introspection. This leads us to the following formal definition of **out-flow**:

$$\text{out-flow}(\mathbf{J}) = \mathbf{J}\mathbf{1} - \text{diag}(\mathbf{J}), \quad (2)$$

where $\mathbf{1}$ is a column vector of ones of length k .

Whereas the rows of the impact matrix \mathbf{J} represent the impact each community has on others, each of the *columns* represent the distribution of impacts other communities have on the community. Therefore, the sum over a column corresponds to the total *in-flow* from other communities to the community. After subtracting the diagonal elements for the same reasons as above, we define **in-flow** as:

$$\text{in-flow}(\mathbf{J}) = \mathbf{J}^T\mathbf{1} - \text{diag}(\mathbf{J}) \quad (3)$$

While some communities may impact a relatively small circle of other communities, others may be broadly influential. Analogously, a community may be influenced by many other communities or it may be strongly influenced just by a few communities. An analysis of the distribution of out-flow (rows of the impact matrix) and in-flow (columns) gives a clear indication of whether a community’s out-flow/in-flow is largely dispersed or narrowly focused. We quantify the heterogeneity of out-flow or in-flow as an entropy of a row or a column of the impact matrix \mathbf{J} . Because some elements of \mathbf{J} may be 0, let us use the convention $\log_2(0) = 0$. Furthermore it is necessary to normalise the rows of the matrix in order to obtain probability distributions of impact, i.e. $\mathbf{J}_{uv}^N = \mathbf{J}_{uv} / \sum_{x=1}^k \mathbf{J}_{ux}$. Formally, we define normalised **out-flow entropy** as:

$$\text{ent}_{\text{of}}(u, \mathbf{J}) = - \frac{\sum_{x=1}^k \mathbf{J}_{ux}^N \log_2 \mathbf{J}_{ux}^N}{\log_2 k} \quad (4)$$

Analogously, the normalised **in-flow entropy** is defined similarly but on the transpose \mathbf{J}^T :

$$\text{ent}_{\text{if}}(u, \mathbf{J}) = \text{ent}_{\text{of}}(u, \mathbf{J}^T) \quad (5)$$

Both measures range within $[0, 1]$. The more the out-flow (in-flow) of community u is equally distributed, the more

the entropy value is close to 1. We note that in the case of entropy we *include* the diagonal elements (introspections), because we wish to differentiate whether the most of the community’s total impact is concentrated *within* that community or not.

Data

Our analysis focuses on AI communities as we are familiar with the main paradigms and events within it. We adopted Martins et al.’s (Martins et al. 2010) classification of computer science conferences (Perfil-CC) into categories A, B or C according to their presumed merit. We further merged the Perfil-CC category Machine Learning (ML) with the category AI. This results in 87 different AI conferences in total.

Our dataset is based on merged records from Arnet-Miner (Tang et al. 2008) and CiteSeerX (Li et al. 2006) from September 2013. The full description of the preparation of the dataset and the data itself is available in the online supplementary material. We analysed a subset of citation data between 1990 and 2008 due to a low number of records prior and after those years. Out of the 87 AI conferences listed by Perfil-CC, we found 59 in our dataset. This corresponds to a coverage of nearly 68%. We segment the data using a sliding time-window. As some of the conferences that we analysed are held biennially (e.g. IJCAI), we chose a sliding window of 4 years. In order to investigate the main trends of the aggregate measures, we divided the entire *analysis period* into 3 sub-periods: *early* period between the years 1990–1996; *middle* period between 1997–2002; and *late* period covering the years 2003–2008.

Inter-Community Analysis of AI Conferences

The following are our exploratory hypotheses:

High *out-flow* (Equation 2) and *introspection* (self-impact) indicate a strong community, with strong internal discourse (introspection) and strong consumption by the other communities (out-flow). Following (Goldstone and Leydesdorff 2006) we may say that such communities are net “exporters”.

High *in-flow* (Equation 3) and low *introspection* indicates a community that acts as a *hub* bringing together researchers from diverse communities. Since those researchers are likely to cite the papers published in their core communities, we expect a high in-flow to the hub community.

A community that is growing increasingly *introspective* and isolated (low *out-flow* and *in-flow*) may be a community currently in relative decline, as these figures suggests that the community is not importing external ideas, nor is it exporting.

Figure 3 depicts the mean *out-flow* (x-axis) and *introspection* (y-axis) of the AI communities in each period. For the sake of brevity, we only discuss six communities, selected because the values of one or more of their aggregate measures suggest that they represent characteristic examples of the types of communities that we described above, i.e. hubs, exporters, or self-referential communities. In the online supplementary material, the reader can interactively explore

these trends. In Table 1 we list the mean values of the aggregate measures for each community along with the community size.

COLT (A) In the early period, we see that the community COLT (ANNUAL CONFERENCE ON COMPUTATIONAL LEARNING THEORY) had both very high introspection and out-flow. This suggests that the community was strong as it maintained high level of both internal discourse but at the same time its outcomes were referenced from the outside. Over the time, the out-flow of COLT has increased substantially while its introspection lowered. Together with the fact that Perfil-CC ranked COLT as a class A conference, it indicates that COLT has evolved from a relatively highly self-referential community into a more open community, while it has increased its already high impact on the other communities.

NIPS (A) Another community with a relatively high introspection in the early period was NIPS (CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS). While maintaining high level of internal discourse over the time, the out-flow from NIPS has increased considerably in the middle and especially in the late period. As its name suggests, NIPS started as a conference with predominantly computational neuroscience focus. Over the time, however, it became one of the major venues in machine learning, artificial intelligence, and statistics. This transition to a more open conference with a broader focus is also indicated by a rise of the community’s in-flow (see Table 1). We may therefore conjecture, that the increase of its total impact can be attributed to the successful transition from a tight community with a narrow focus to a more open community with strong level of internal discourse.

community	OF	INT	INT/OF	IF	size	period
CBR	0.08	0.13	1.64	0.39	88	early
CBR	0.47	0.83	1.74	0.91	152	middle
CBR	0.69	0.56	0.81	1.07	119	late
COLT	0.87	1.00	1.15	1.86	71	early
COLT	1.61	0.91	0.57	1.15	48	middle
COLT	2.04	0.44	0.21	0.60	55	late
ICML	0.42	0.08	0.18	0.65	66	early
ICML	1.76	0.24	0.13	0.96	139	middle
ICML	3.65	0.64	0.18	3.12	210	late
IJCAI	0.44	0.07	0.16	5.60	305	early
IJCAI	1.22	0.11	0.09	4.60	231	middle
IJCAI	1.51	0.09	0.06	7.01	282	late
ILP	0.92	0.75	0.82	0.23	38	middle
ILP	0.44	0.36	0.82	0.68	51	late
NIPS	0.17	0.16	0.97	0.03	537	early
NIPS	0.57	0.29	0.50	1.08	447	middle
NIPS	1.12	0.34	0.30	2.55	498	late

Table 1: Mean aggregate measures out-flow (OF), introspection (INT), in-flow (IF), size, and the ratio of introspection and out-flow (INT/OF) of the AI conferences that are discussed in the main text. The figures mentioned in the text are in bold. The figures are rounded for the sake of brevity.

IJCAI (A) One of the most respected conferences in AI is the INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE (IJCAI) that has been held biennially since 1969. As a rather large conference with a broad focus covering many sub-disciplines of artificial intelligence, IJCAI is

a typical *hub* venue, where many researchers and practitioners from various fields and of various background meet. It is therefore no surprise that it is characterised by relatively small introspection, but very high in-flow (see Table 1). This indicates that IJCAI attracts researchers, who publish frequently in other communities with a perhaps narrower focus that corresponds to their domain of expertise, and who seek to disseminate the results of their work beyond the boundaries of their core communities.

ICML (A) ICML (INTERNATIONAL CONFERENCE ON MACHINE LEARNING) is a premium machine learning conference (class A in Perfil-CC). In contrast with IJCAI, its higher introspection suggests that it is less of a hub and that its attendees tend to regularly publish their work at it. The more than 8 times increase of its out-flow between the early and late periods suggests a rising interest in and consumption of the machine learning methods within other communities. ICML thus became an “exporter” of the machine learning techniques.

CBR (B) While both the NIPS and CASE-BASED REASONING (CBR) communities had similar levels of introspection early in the period, CBR remained highly introspective (i.e. self-referential) also in the middle and late periods. A similar trend of a high introspection relative to the out-flow can be observed also for the ILP community (INTERNATIONAL CONFERENCE ON INDUCTIVE LOGIC PROGRAMMING). A high ratio of introspection to out-flow (see Table 1) indicates a community with robust internal discourse but low levels of engagement beyond its immediate borders. While not necessarily a negative indicator, the community may face challenges due of its relative lack of visibility. In the analysis period, we observe that the size of the CBR community grew from 88 in the early period to 152, but then it lowered to 119 in the later period. In order to shed some light on these trends, in the next section we take CBR as a subject of a more analysis.

In-depth analysis of the CBR community

We validate our findings by a qualitative analysis of the history of the CBR community and by contrasting the observed trends of the COIN measures with other measures frequently occurring in the literature. In order to refer to time in a consistent way, we refer to a time-window $[t, t']$ only by its end year t' . For example, the values for the year 1996 were measured in the window $[1993, 1996]$.

CBR: In-flow, Out-flow and Introspection

Figure 4a depicts the change of in-flow, out-flow and introspection measures over the analysis period. From its inception, the community’s introspection steadily increased up to the year 2001, when it peaked and subsequently declined to a stable level between the years 2002–2008. The community’s in-flow reached its peak in 2002, then experienced negative growth until the year 2006, when, however, the trend reversed again. The community had tended to have

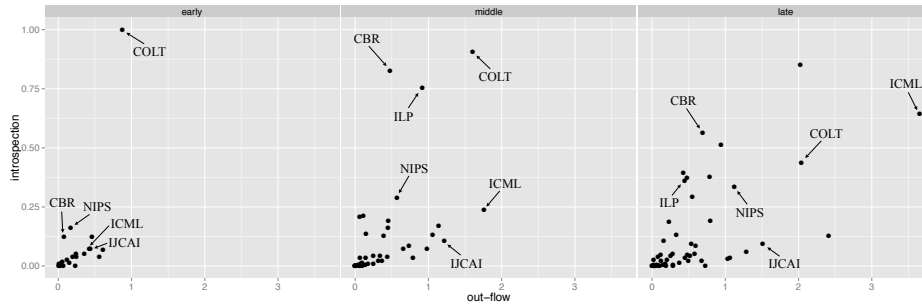


Figure 3: The mean out-flow and introspection for each AI community in early, middle, and late periods. For the sake of brevity, only the communities that are discussed in the main text are annotated. Please note that ICML was very close to IJCAI in the early period because we measured very similar introspection and out-flow for both of them. ILP is not depicted in the early period, because it appears in ArnetCite for the first time in 1997.

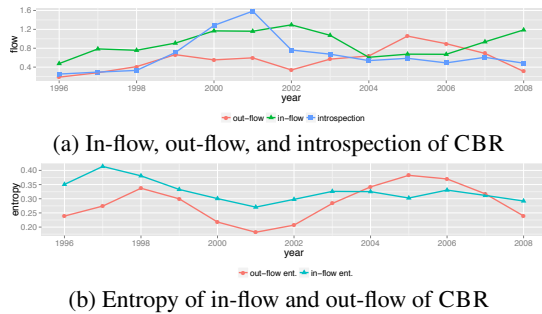


Figure 4: In-flow, out-flow, and their heterogeneity (entropy) for CBR.

relatively low out-flow, except noticeably in 2005. This suggests that while CBR imported from other communities, its exports were often much lower.

As a basis of comparison in the same time period, we provide plots for two conferences : a class A conference, NIPS (Figure 5a), and a class B conference, JELIA (EUROPEAN CONFERENCE ON LOGICS IN ARTIFICIAL INTELLIGENCE) (Figure 5b). In both cases the introspection is low relative to in-flow and out-flow. Even though the in-flow of NIPS is higher than its out-flow, both figures are growing and thus may be interpreted an indicator of the openness of this community. In contrast, the high introspection of CBR and its low out-flow (relative to the two other communities) suggests a community with less engagement with other communities. The out-flow of CBR reaches its peak value in 2005 and then it begins to fall rather sharply. Possibly, the high out-flow in 2005 was induced by the research outputs from the beginning of the time-window [2002,2005], which is close to the peak in introspection measured in 2001. Without further observations of similar trends in other communities, we must speculate that the strong discourse of the community around 2001 translated into impact on other communities.

CBR: Relationship Focus Figure 4b illustrates the change of entropy, i. e. heterogeneity, of the in-flow and out-flow of CBR. In the very beginning, the CBR was influenced by many other communities. For a young paradigm,

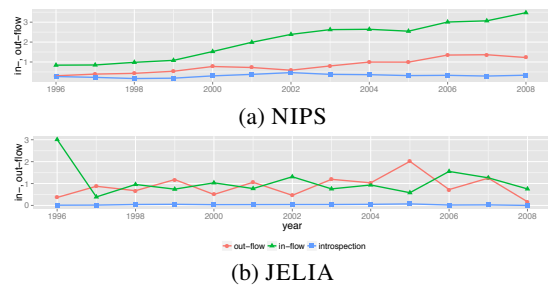


Figure 5: In-flow, out-flow, and introspection for NIPS and JELIA.

this might be expected as it is still yet to develop its own discourse. Since the peak in 1997, its in-flow entropy was steadily decreasing until 2001. Similarly to that, the heterogeneity of its out-flow was increasing at first, but then the out-flow entropy reached its minimum in 2001. The dips in 2001 match the very high introspection in that period, suggesting that the majority of citation activity was fuelled by internal discourse. After that, both in-flow and out-flow entropy increased until 2005–2006, after which they have fallen, suggesting a gradual narrowing focus of the community in the last years of our data. In order to compare and validate our observations, we provide a number of alternative measures of the CBR community.

CBR: Membership Dynamics In order to understand the high levels of introspection of CBR, we examined how stable its core community of members were. Our hypothesis is that greater introspection means less change in the underlying membership. We quantified the stability of the member base of a community as its self-similarity measured by Jaccard index. In our experiment, we compared the stability of the CBR member base with the member base of communities that are also classified as class B conferences. For each time-window between 1993–2008, we measured for each AI community u a Jaccard similarity of the fuzzy sets of its members in time-window t and subsequent time-window

$t + 1$:

$$js(\mathbf{M}_{\cdot u}^t, \mathbf{M}_{\cdot u}^{t+1}) = \frac{\sum_x^n \min(\mathbf{M}_{xu}^t, \mathbf{M}_{xu}^{t+1})}{\sum_x^n \max(\mathbf{M}_{xu}^t, \mathbf{M}_{xu}^{t+1})}, \quad (6)$$

where $\mathbf{M}_{\cdot u}^t$ represents the fuzzy set of the members of community u at time-window t (see Section).

We computed an expected value of the Jaccard similarity for each window excluding the values of CBR. This way we obtained a paired sample of two time-series: one for the “average” class B conference and one for the CBR community itself. We found that CBR had significantly higher self-similarity than the rest of the class-B communities (Wilcoxon signed-rank test, $p = 0.02$).

CBR: Google Trends and Publication Output Using the Google Trends service we examined the number of queries for case-based reasoning from January 2004 to September 2013² From a peak in March 2004 of 100 searches, the number has diminished to about 10 searches per month from 2009 onwards. This fall coincides with the drop in out-flow illustrated by Figure 4a. We also looked at the number of regular papers published by CBR per year. The highest number of papers was in the years 1997 (60 papers) and 2004 (59 papers). Then it lowered to 35–43 between 2006–2012.

CBR: Other Performance Measures We investigated three additional statistics per time-window: PageRank (PR), group in-degree (GI) (Everett and Borgatti 1999) and 3-year conference impact factor (CIF) (Martins et al. 2010). Group in-degree is the total number of citations received by the papers published by CBR from papers published elsewhere. Conference impact factor (CIF) of a conference in year t is the average number of citations a paper published by the community within $[t - 3, t - 1]$ received from all the papers published at t . All measures except CIF were decreasing since 2005. While GI was falling since 2005, we observed a rather moderate rise of CIF since that year. Since CIF includes self-citations whereas GI does not, the rise of CIF may be attributed to the rising introspection of CBR.

Conclusion

Our analysis reveals how the relationship between scientific communities may evolve and the different roles a community may play. In the context of the CBR community our analysis contrasts the finding of a purely introspective analysis carried out by the CBR community itself in 2008 (Greene et al. 2008), in which the authors suggested that regular rise of new topics within the community “can be considered a sign of a healthy research area” (Greene et al. 2008). We would suggest that every community needs to examine its external as well as internal relationships before drawing such a conclusion.

²See <https://www.google.com/trends/explore?q=case+based+reasoning>. The trend using the “case-based reasoning” phrase looks similar. Data prior 2004 are unavailable due to the limitations of the service.

Acknowledgements

This research has been funded by Science Foundation Ireland under Grant Numbers SFI/08/SRC/I1407 (Cliques) and SFI/12/RC/2289 (Insight).

References

- Belák, V.; Karnstedt, M.; and Hayes, C. 2011. Life-cycles and mutual effects of scientific communities. *Procedia - Social and Behavioral Sciences* 22:37–48.
- Belák, V.; Lam, S.; and Hayes, C. 2012. Cross-community influence in discussion fora. In *Proceedings of 6th International Conference on Weblogs and Social Media*. AAAI.
- Biryukov, M., and Dong, C. 2010. Analysis of computer science communities based on DBLP. In *Research and advanced technology for digital libraries*. Springer. 228–235.
- Dietz, L.; Bickel, S.; and Scheffer, T. 2007. Unsupervised prediction of citation influences. In *Proceedings of ICML’07*. ACM.
- Everett, M. G., and Borgatti, S. P. 1999. The centrality of groups and classes. *Journal of Mathematical Sociology* 23(3):181–201.
- Goldstone, R. L., and Leydesdorff, L. 2006. The import and export of cognitive science. *Cognitive Science* 30(6):983–993.
- Greene, D.; Freyne, J.; Smyth, B.; and Cunningham, P. 2008. An analysis of research themes in the CBR conference literature. In *Advances in Case-Based Reasoning*. Springer.
- Jones, B. F.; Wuchty, S.; and Uzzi, B. 2008. Multi-university research teams: shifting impact, geography, and stratification in science. *Science* 322(5905):1259–1262.
- Li, H.; Councill, I.; Lee, W.-C.; and Giles, C. L. 2006. CiteSeerX: an architecture and web service design for an academic document search engine. In *Proceedings of WWW’06*, 883–884. ACM.
- Martin, B. R., and Irvine, J. 1983. Assessing basic research: some partial indicators of scientific progress in radio astronomy. *Research policy* 12(2):61–90.
- Martins, W. S.; Gonçalves, M. A.; Laender, A. H.; and Ziviani, N. 2010. Assessing the quality of scientific conferences based on bibliographic citations. *Scientometrics* 83(1):133–155.
- Moed, H. F. 2005. *Citation analysis in research evaluation*, volume 9. Springer.
- Montolio, S. L.; Dominguez-Sal, D.; and Larriba-Pey, J. L. 2013. Research Endogamy as an Indicator of Conference Quality. *SIGMOD Record* 42(2):11.
- Patil, A.; Liu, J.; and Gao, J. 2013. Predicting group stability in online social networks. In *Proceedings of the international conference on World Wide Web (WWW’13)*. ACM.
- Tang, J.; Zhang, J.; Yao, L.; Li, J.; Zhang, L.; and Su, Z. 2008. ArnetMiner: extraction and mining of academic social networks. In *Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining*.